Network mechanisms regulating the generation of sharp wave-ripple complexes in the hippocampus

The role of disinhibition in three-population models

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

M.Sc. Roberta Evangelista

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Bernhard Grimm

Gutachter/innen:

1. Prof. Dr. Richard Kempter
2. Prof. Dr. Dietmar Schmitz
3. Prof. Dr. Christian Leibold

Tag der mündlichen Prüfung: 20. September 2019

Abstract

Sharp wave-ripple complexes (SWRs) are brief (50-100 ms) events of coordinated and increased network activity originating in the CA3 region of the hippocampus. SWRs are thought to mediate the consolidation of explicit memories, but the mechanisms underlying their occurrence remain obscure. Characteristic of SWRs is that they modulate the firing of different classes of hippocampal cells. Pyramidal cells (PYR) and parvalbumin-positive basket cells (PV^+BCs) preferentially fire during SWRs and are almost silent outside. SWRs emerge spontaneously *in vivo* and *in vitro*, but can also be triggered by activating PYR, and, interestingly, by activating PV^+ cells. How can the activation of PV^+ interneurons result in an increase of PYR firing? To answer this question, I explore how the interaction of excitatory neurons (PYR) and two groups of competing interneurons (PV^+BCs and a class of yet unidentified anti-SWR cells) contributes to the initiation, termination, and incidence of SWRs.

In the first part of this thesis, I introduce a biophysically constrained network of spiking neurons and show that it can exhibit spontaneous, fluctuation-driven transitions from a non-SWR state (dominated by active anti-SWR cells) to a SWR state, in which active PV^+BCs disinhibit PYR by suppressing anti-SWR cells. Additionally, SWR events can be triggered by activating PYR or PV^+BCs, or inactivating anti-SWR cells. Short-term synaptic depression at the PV^+BCs-to-anti-SWR cells connections regulates the termination of SWR events.

The coexistence of non-SWR and SWR states for intermediate values of the synaptic depression allows to study the network behavior in terms of bistability. In the second part of this thesis, I reduce the spiking network to a mean-field approximation, where conditions for the emergence of a bistable configuration are derived analytically. The mean-field approximation strongly decreases the model complexity and allows to unveil the mechanisms regulating the existence of this type of bistable disinhibitory networks.

The model predicts the existence of a class of anti-SWR cells that are tonically active in the non-SWR state and silent during SWR events. In the last part of this work, I show the first experimental evidence for CA3 interneurons that are anti-modulated with respect to SWRs, and discuss their potentially causal involvement in the SWR generation process.

Overall, the results of this thesis elucidate the role of interneurons in SWR generation and broaden our understanding of the microcircuits supporting the dynamics of memory-related networks.

Zusammenfassung

Sharp wave-ripple Komplexe (SWRs) sind kurze (50-100 ms) Ereignisse von erhöhter und kohärenter Netzwerkaktivität. Es wird angenommen, dass SWRs, die der CA3-Region des Hippocampus entspringen, eine wichtige Rolle bei der Konsolidierung von expliziten Gedächtnisinhalten spielen. Allerdings sind die SWRs zugrundeliegenden Mechanismen bis heute ungeklärt. Charakteristisch für SWRs ist, dass sie die Aktivitäten von verschiedenen Neuronenklassen modulieren. Pyramidenzellen (PYR) und Parvalbumin-positive Korbzellen (PV^+BCs) feuern während SWRs besonders häufig, wohingegen sie außerhalb beinahe inaktiv sind. SWRs treten spontan *in vivo* und *in vitro* auf, können aber auch durch Stimulation von PYR und interessanterweise auch von PV^+ Zellen hervorgerufen werden. Wie kann die Stimulation dieser inhibitorischen Interneuronen zu einer erhöhten Aktivität von PYR führen? Um diese Frage zu klären, untersuche ich wie das Zusammenspiel von exzitatorischen Neuronen (PYR) und zwei Klassen von konkurrierenden Interneuronen (PV^+BCs und derzeit unbekannte Anti-SWR-Zellen) sowohl die Entstehung von SWRs beeinflusst als auch die Dauer und die Häufigkeit dieser Ereignisse steuert.

Im ersten Teil dieser Arbeit entwickle ich, basierend auf experimentellen Randbedingungen, ein Netzwerk aus feuernden Neuronen. Dieses Netzwerk zeigt spontane, fluktuationsinduzierte Übergänge vom Anti-SWR-Zustand zum SWR-Zustand. Der Anti-SWR-Zustand wird von aktiven Anti-SWR-Zellen dominiert; der SWR-Zustand wird dominiert von der Aktivität von PV^+BCs, die Anti-SWR-Zellen inhibieren und somit die Aktivität von PYR disinhibieren. Außerdem können SWRs hervorgerufen werden sowohl durch Stimulation von PYR oder PV^+BCs als auch durch Inaktivierung von Anti-SWR-Zellen. Durch einen Kurzzeitdepressionsmechanismus der synaptischen Verbindung von PV^+BCs zu Anti-SWR-Zellen wird die Dauer der SWRs reguliert.

Die Koexistenz von Anti-SWR- und SWR-Zuständen bei mittleren, konstanten Stärken der synaptischen Depression erlaubt die Untersuchung der Bistabilität des Netzwerks. Im zweiten Teil der Arbeit reduziere ich die Modellkomplexität des Netzwerkes durch eine Mean-field-Näherung, sodass Voraussetzungen für bistabile Netzwerkaktivität analytisch hergeleitet werden können. Dies erlaubt es, Mechanismen zu identifizieren, die die Existenz der Bistabilität von disinhibitorischen Netzwerken dieser Art beeinflussen.

Das Modell prognostiziert die Existenz einer Klasse von Anti-SWR-Zellen, die im Anti-SWR-Zustand tonisch aktiv und während des SWR-Zustandes inaktiv sind. Im letzten Teil dieser Arbeit zeige ich erste experimentelle Ergebnisse, die die Existenz von CA3-Interneuronen belegen, die anti-moduliert sind bezüglich SWRs. Weiterhin diskutiere ich deren möglichen kausalen Einfluss auf SWR-Generierung.

Durch Untersuchung der Rolle von Interneuronen hinsichtlich der Generierung von SWRs trägt diese Arbeit zu einem tieferen Verständnis der neuronalen Schaltkreise im Hippocampus bei, die essentiell für den Erwerb und die Konsolidierung explizieter Gedächtnisinhalte sind.

Contents

Abbreviations

AAC	Axo-Axonic Cell
ACh	Acetylcholine
aCSF	Artificial Cerebrospinal Fluid
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic, receptor
CA	Cornu Ammonis (with subfields CA1, CA2, CA3)
CCK$^+$	Cholecystokinin-positive, cell
DG	Dentate Gyrus
EC	Entorhinal Cortex
EEG	Electroencephalogram
ENK	Enkephalin-expressing, cell
FP	Fixed Point
FWHM	Full Width at Half Maximum
GABA	γ-Aminobutyric acid, neurotransmitter and receptor
IEI	Inter-Event-Interval
LIF	Leaky Integrate-and-Fire, neuron model
LFP	Local Field Potential
LTD	Long-Term Depression
LTP	Long-Term Potentiation
MTL	Medial Temporal Lobe
NMDA	N-methyl-D-aspartate, receptor
ODE	Ordinary Differential Equation
PSC	Postsynaptic Current (excitatory, EPSC, or inhibitory, IPSC)
PSP	Postsynaptic Potential (excitatory, EPSP, or inhibitory, IPSP)
PTI	Perisomatic-Targeting Interneuron
PV$^+$BC	Parvalbumin-positive Basket Cell
SD	Standard Deviation
SEM	Standard Error of the Mean
SOM$^+$	Somatostatin-positive, cell
STDP	Spike-Timing Dependent Plasticity
SWR	Sharp Wave-Ripple complex
SW	Sharp-Wave (low-pass filtered component of a SWR)
SWS	Slow-Wave Sleep

1 Introduction

1.1 Memory and the two-stage model

The ability to store and recall memories constitutes the essence of our human nature. It is the collections of experiences, learned skills and semantic knowledge that allows us to communicate, act and evolve. As such, the question of how memory systems work remains a fundamental issue in modern neuroscience.

Even though not much was known about the mechanisms underlying memory formation, the concept that newly created memories undergo stabilization and maturation was already present in the writings of Quintilian, a Roman educator and rhetorician from the first century A.D. He noticed 'a curious fact... that the interval of a single night will greatly increase the strength of the memory', and suggested that 'the power of recollection... undergoes a process of ripening and maturation during the time which intervenes' [Dudai, 2004]. The process known today as memory consolidation (or Konsolidierung, [Müller and Pilzecker, 1990]), defined as 'the progressive post-acquisition stabilization of memories' [Dudai, 2004], is believed to follow memory acquisition, a stage (also known as memory encoding) during which new memories are formed. Classically, memory consolidation is divided in two families of processes: one occurring for minutes to hours after the encoding (synaptic or local consolidation), followed by a process lasting weeks, months or even years (system consolidation), that is believed to involve the reorganization of memories over time [Squire and Alvarez, 1995; Frankland and Bontempi, 2005].

For centuries, philosophers and anatomists have struggled to identify the location in the brain where memories are stored [Martinez and Kesner, 2014]. A significant breakthrough came in the late 1950's from the study of H.M. [Scoville and Milner, 1957], an epileptic patient whose medial temporal lobes (MTLs) were removed to ameliorate his seizures. At the time, it was known that epileptic episodes originate in the MTL area, but it was surprising to discover that the patient started suffering from severe memory deficits after the surgery was performed. H. M. lost the ability to create new memories (*anterograde amnesia*), and was unable to recall episodes that occurred up to three years before the surgery. However, he could still recall episodes from his childhood (graded *retrograde amnesia*), and perform and learn new tasks. This case study was instrumental in two ways. First, it provided the first evidence for the existence of a region responsible for the formation (and, at least in part, for storage and recall) of *episodic memory* (the

memory for events in a spatio-temporal context). Second, the observation that episodic memory was strongly impaired, whereas *procedural memory* (the type of memory describing the ability to learn motor skills) was unaffected, suggested that these two processes involve different brain structures. Nowadays, the hippocampus (which is part of the MTL) is recognized as the hub for the encoding of episodic memory, and is thought to be involved in the recall of already stored memories [Squire, 1992; Eichenbaum, 2000]. On the other hand, procedural memory is believed to rely on the cerebellum and striatal cortex [Gabrieli, 1998]. In this thesis, I primarily focus on the processes underlying the consolidation of episodic memory.

This chapter is organized as follows. In Section 1.1.1, I introduce the concept of synaptic plasticity, which is believed to constitute the biological substrate of memory. Moreover, I present the two-stage model, a theory that hypothesizes that episodic memory acquisition and consolidation are distinct processes mediated by different brain states. The hippocampus and other cortical regions are believed to be the most important areas involved in these processes. Thus, I briefly review the cortico-hippocampal circuitry in Sec. 1.1.2. Section 1.1.3 focuses on hippocampal rhythms occurring during active and offline behavioral states, and presents experimental evidence supporting the involvement of offline hippocampal rhythms (the sharp wave-ripple complexes, or SWRs) in the consolidation process. In Sec. 1.1.4, I introduce the main corresponding rhythms in cortical structures, and discuss their contribution to the consolidation process and their relation to hippocampal activity. Furthermore, Section 1.2 addresses the question of SWR generation: here I present existing hypotheses and models on the mechanisms regulating the emergence of SWRs, and I review the main experimental features that the model I am proposing should incorporate.

1.1.1 The biological basis of memory

From a biological perspective, the question of how memories are encoded in the brain remains elusive. One of the first hypotheses on the neural representation of memory was formulated by Richard Semon [Semon, 1921], who suggested that learning activates a small ensemble of neurons and induces persistent changes of their connections. This defines an *engram*, a biological trace of a memory, whose reactivation results in the recall of the memory itself. To explain how the persistent changes postulated by Semon could arise, Donald Hebb proposed *synaptic plasticity* as a candidate mechanism for memory encoding [Hebb, 1949]. The concept of synaptic plasticity refers to activity-dependent changes in the strength of the connections among neurons: when the pre- and postsynaptic neurons are simultaneously active, the connecting synapse is strengthened. Thus, an engram emerges as an association of active cells with strengthened connections (*cell assembly*).

Two decades after Hebb's formulation, synaptic plasticity mechanisms known as *long-term potentiation* (LTP) and *long-term depression* (LTD) had been observed throughout

the hippocampal formation [Bliss and Lomo, 1973; Lynch et al., 1977; Martin et al., 2000; Whitlock et al., 2006; Neves et al., 2008], and are nowadays accepted to be the biological basis of memory formation. The dissection of the cellular mechanisms underlying LTP and LTD revealed the existence of an early and a late phase of plasticity [Govindarajan et al., 2006; Mayford et al., 2012]. During early LTP (e-LTP), NMDA-receptors mediate an increase in intracellular Ca^{2+} concentration, that transiently increases the conductance at AMPA-receptors. Additionally, this potentiation is supported by an increase in the number of dendritic spines. During late LTP (l-LTP), which lasts for several hours after encoding, the changes in synaptic weights are maintained by protein synthesis-dependent processes; this phase roughly corresponds to the phase of synaptic consolidation defined above. Despite these exciting discoveries, it remains to be explained how a complex episodic memory (as, for example, the memory of your first day of school) is distributed and encoded at the level of the synapses, and how the different memory 'pieces' stored in multiple synapses can be combined to reconstruct the complete memory. Recently, the lab of Susumu Tonegawa performed a series of optogenetic experiments to identify memory engrams in rodents. By combining activity-dependent cell labeling and optogenetics, the researchers could identify cells belonging to the engram of a newly induced fear memory [Liu et al., 2012]; the activation of these cells induced memory retrieval. In a follow-up study, Ryan et al. [2015] showed that the synaptic connections of active engram cells revealed traces of LTP (increased synaptic strength and larger spine density, see Tonegawa et al. [2018] for a review). These promising results shed some light onto the fascinating question of 'what is the engram?', which remains at the core of memory research [Poo et al., 2016].

The fact that H.M. could still recall remote memories suggests that after a certain period of time memories can become independent of the MTL while possibly persisting in other areas. A potential candidate is the neocortex [Squire and Alvarez, 1995; Frankland and Bontempi, 2005; Neves et al., 2008], an area that is connected to the hippocampus via the entorhinal cortex (EC) and is known to be less plastic than the hippocampus [Castro-Alamancos and Connors, 2002; Feldman, 2009]. This feature is particularly advantageous to prevent deleterious forgetting caused by new incoming memories, which could overwrite existing memories if the system was too plastic. The existence of the neocortex as a stable location for memory storage constitutes a possibility to overcome the so called *plasticity-stability dilemma* [Abraham and Robins, 2005]. Interestingly, the coexistence of different memory systems has been shown to be advantageous also from a computational perspective, because it allows for complementary processes (memory acquisition and consolidation) to happen at separate locations and with different timescales [Marr, 1971; McClelland et al., 1995; Káli and Dayan, 2004]. The degree at which long-term stored memories become hippocampus-independent is still a matter of debate [McClelland et al., 1995; Nadel and Moscovitch, 1997; Frankland and Bontempi, 2005; Winocur et al., 2010], and might ultimately depend on multiple factors, such as memory type and content, reactivation, etc.

Two-stage model. To answer the question of how memories could be stored and retrieved, György Buzsáki formalized the *two-stage model* of memory trace formation [Buzsáki, 1989]. According to this model, memories are formed during an exploration phase, where the animal receives information from the external world. The sensory inputs arrive at cortical areas and are then transmitted to the hippocampus. During this phase, synaptic plasticity is used to store events as labile memory traces. Successively, these newly encoded traces are reactivated during the offline state of memory consolidation. The reactivation, or *replay*, of stored memories is the key feature needed to induce long-term synaptic modifications, which allows memories to become long-lasting and avoid erasure over the following exploration phases. The model intends to overcome the known limitation of synaptic plasticity decay over time [Huang, 1998; Abraham and Williams, 2003] by proposing a reactivation mechanism of the cells encoding a memory during a state in which sensory inputs are largely absent. Hence, this model proposes a crucial role for offline states, e.g. sleep states, for the successful consolidation of memory traces. This hypothesis is in line with longstanding research on the beneficial effect of sleep on memory performance [Marshall and Born, 2007; Diekelmann and Born, 2010].

A key feature of the two-stage model is that it associates the two-stages of memory encoding and consolidation with behaviorally observed hippocampal rhythms. Before we discuss these phenomena in detail, it is useful to review a few key concepts about cortico-hippocampal anatomy.

1.1.2 Brief overview of cortico-hippocampal anatomy

The hippocampus is a C-shaped structure located in the medial temporal lobe. Mammals have two hippocampi, one per hemisphere, and these structures have been shown to be remarkably similar in different species [Manns and Eichenbaum, 2006; Clark and Squire, 2013]. The hippocampus comprises the Cornu Ammonis regions (CA3, CA2, CA1), the dentate gyrus (DG) and the subiculum. A peculiar feature of both the Cornu Ammonis and the DG is their laminar structure. The cell bodies of the principal cells of the DG (the granule cells) are organized in a dense layer; similarly, the pyramidal cells in CA1, CA2, CA3 have cell bodies in the *stratum pyramidale*, with basal dendrites and axons branching out to *stratum oriens*, and apical dendrites extending through *stratum radiatum* and *stratum lacunosum moleculare*. Interneurons, in contrast, have a more variable spatial distribution [Freund and Buzsáki, 1996; Klausberger and Somogyi, 2008]. For example, parvalbumin positive (PV^+) basket cells, a prominent class of hippocampal interneurons, have cell bodies in *stratum pyramidale* or *radiatum*, a dense axonal arbor in *stratum pyramidale* (possibly extending to *strata oriens* and *radiatum*) and dendrites spanning almost all layers [Booker and Vida, 2018].

Inside the hippocampus, information is streamed from the DG to CA3 via the mossy fibers (unmyelinated granule cells' axons), from CA3 to CA1 via the Schaffer collaterals (axons of CA3 cells), and from CA1 to subiculum.

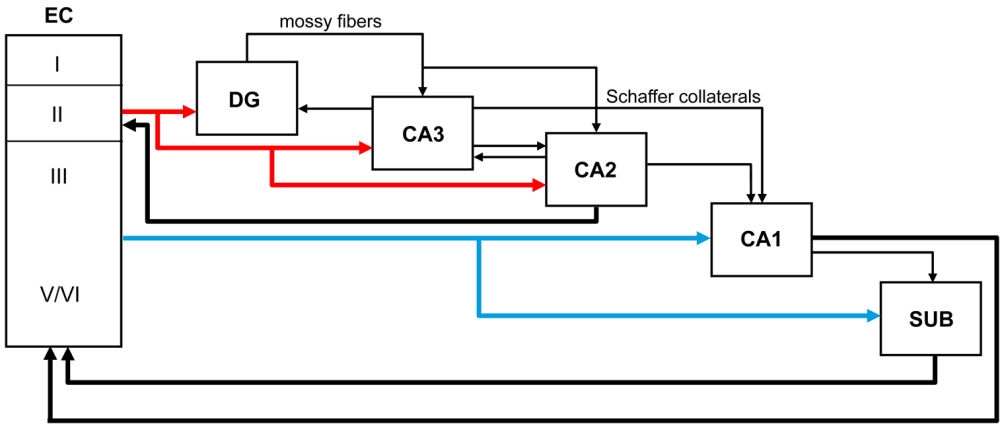

FIGURE 1.1: **Schematic view of cortico-hippocampal connections.** The diagram shows the main excitatory connections between the entorhinal cortex (EC), the dentate gyrus (DG), the CA fields, and the subiculum (SUB). Adapted from Llorens-Martin et al. [2014].

The hippocampus is connected to cortical areas mostly via the entorhinal cortex, which provides both afferent and efferent connections. As the hippocampus itself, the EC is a layered structure, with pyramidal cell bodies distributed in layers II-VI and interneurons variably distributed across all layers (I-VI). The information from superficial layers (I-III) enters the hippocampus via the perforant path, and successively spreads to the DG, CA regions and subiculum. From the subiculum, the information is transmitted back to the deep layers of the EC. Next to this main direction of signal transmission (EC superficial layers → DG → CA3 → CA1 → subiculum → EC deep layers), there are multiple other shortcuts and feedback loops, e.g., a direct EC layer II to CA3 and layer III to CA1 connections, and recurrent connections in CA3 [MacVicar and Dudek, 1980; Witter, 1993; Canto et al., 2008]. Figure 1.1 gives a schematic overview of the existing connections.

Connections between the EC and neocortical areas are widespread [Canto et al., 2008], including connections with prefrontal, piriform, parietal and retrosplenial cortices. Finally, the EC connects to subcortical areas like the amygdala, thalamus, basal ganglia, basal forebrain and brainstem [Burwell and Witter, 2002].

1.1.3 Hippocampal rhythms

To monitor the hippocampal activity across different behavioral states, researchers measure the local field potential (LFP), a signature of the extracellular brain activity. As such, the signal is widely believed to describe the coordinated activity of large neuronal ensembles, even though the details about how the LFP is generated are a matter of intense debate (see Einevoll et al. [2013] and Sec. 2.8.4).

1.1.3.1 Theta oscillations during exploratory behavior

During exploratory periods, the hippocampal LFP of mammals shows oscillations in the theta range (8-12 Hz), [Arnolds et al., 1980; Bland, 1986; Buzsáki, 2002]. These oscillations are complemented by gamma oscillations (slow gamma: 30-50 Hz, fast gamma: 60-90 Hz), nested at every theta cycle [Buzsáki et al., 1983; Bragin et al., 1995a] (see Colgin [2015] for the relevance of theta-gamma coupling for memory). Additionally, the membrane potential of neurons has been shown to oscillate at theta frequency [Alonso and Klink, 1993; Kamondi et al., 1998; Leung and Yu, 2017; Domnisoru et al., 2013]. A very interesting feature is that the oscillation phase at which hippocampal cells fire advances over time, a phenomenon known as phase precession [O'Keefe and Recce, 1993; Skaggs et al., 1996]. If cells that encode a memory fire in sequence at a behavioral time scale, phase precession provides an opportunity to compress the sequence inside a theta cycle. Preconditions are that each cell fires multiple times (so that the phase advance has an effect), and that the order of spiking is preserved over cycles. Compressing a spiking sequence in a theta cycle can be advantageous to elicit synaptic plasticity: multiple cells will be active in concert at short temporal distances, a scenario that is thought to favor the strengthening of synapses [Hebb, 1949]. Interestingly, a synaptic plasticity mechanism known as *spike-timing dependent plasticity* (STDP) [Levy and Steward, 1983; Bi and Poo, 1998] has been hypothesized as a tool to preserve the directionality of the sequence. According to the STDP rule, if a presynaptic cell fires shortly ($\sim$10 ms) before a postsynaptic one, the connection is potentiated, whereas the synaptic strength is reduced when the postsynaptic firing precedes the presynaptic one.

One famous example in which the phase precession mechanism can be observed is the case of *place cells*. Place cells, originally discovered by O'Keefe and Dostrovsky [1971], fire when the animal is at a specific location in space. Place cells cover the space uniformly and are thus suited to function as a tool for spatial navigation. As the animal traverses an arena along a given trajectory, multiple place cells will be sequentially active; decoding the neural activity thus provides information about the animal position in real time (see Fig. 1.3). Phase precession has been shown to compress the sequence of active place cells to the timescale of synaptic plasticity [O'Keefe and Recce, 1993; Skaggs et al., 1996], thus allowing place cells to strengthen their synaptic weights in sequence. In this way, the animal could memorize the trajectory to a given location (e.g., the reward location where food is provided). Sequences are not constrained to represent space: another prominent example are time cells, hippocampal neurons that are active during the delay period of a memory task [Pastalkova et al., 2008; MacDonald et al., 2011; Kraus et al., 2013].

1.1.3.2 Sharp wave-ripples complexes during offline states

The LFP signal switches from theta oscillations during states of active exploration to a more irregular activity during consummatory behavior (as grooming, eating, licking), immobility, and slow-wave sleep. In this state, the LFP is characterized by *sharp wave-ripple complexes* (SWRs): large (up to 2 mV), transient deflections (the sharp waves,

SWs), overlaid by fast (ripple) oscillations (∼120-250 Hz). These events can be observed throughout the hippocampus, most prominently in the CA areas, and occur irregularly with an incidence of around 0.5-1/s, and a duration of ∼50-100 ms [Maier and Kempter, 2017]. SWRs show a typical profile across different somato-dendritic layers: the sharp wave has negative polarity in *strata radiatum* and *lacunosum moleculare*, but polarity reverts in *strata pyramidale* and *oriens* [Buzsáki et al., 1983; Buzsáki, 1986; Sullivan et al., 2011]. Negative polarity in *stratum radiatum* is thought to represent the contribution of incoming excitatory inputs from CA3 to the apical dendrites of both CA1 and CA3 itself. Ripple oscillations are observed across strata, but ripple power is strongest in *stratum pyramidale*, suggesting that it might be influenced by the activity of perisomatic interneurons [Buzsáki et al., 1992; Ylinen et al., 1995; Csicsvari et al., 2000].

Even though SWRs were first defined by Buzsáki et al. [1983], these rhythms had already been observed in Jouvet et al. [1959]; Vanderwolf [1969]; Hartse et al. [1979] as activity signatures alternative to theta oscillations observed during exploration. The dichotomy of SWRs and theta oscillations is also present during sleep, with slow-wave sleep (SWS) displaying SWR events and REM sleep characterized by awake-like theta oscillations.

Sharp waves are already visible in rodents at postnatal day 3, and thus constitute the first organized network event in the hippocampus, whereas ripples develop starting from the second week after birth [Leinekugel et al., 2002; Buhl and Buzsáki, 2005]. This finding also hints at the existence of separate mechanisms for the generation of the two components of the SWR. SWRs seem to be conserved across different mammalian species (rodents, cats, primates, humans, bats) [Buzsáki, 2015]. Their existence in non-mammalians is more debated. They have been observed in reptiles [Shein-Idelson et al., 2016] and in zebrafishes [Vargas et al., 2012], but not in birds [Rattenborg et al., 2011].

SWRs constitute the most synchronous event in the mammalian brain [Buzsáki, 2015], and strongly enhance cell excitability in the hippocampus and neighboring structures [Buzsáki, 1986; Chrobak and Buzsáki, 1994; Csicsvari et al., 1999a]. During SWRs, neurons show increased synchrony [Ylinen et al., 1995; Csicsvari et al., 2000], even though only a fraction (0-50%, [Ylinen et al., 1995; Mizuseki and Buzsáki, 2013]) of neurons participates in each event and a participating neuron might spike only once during an event [Csicsvari et al., 1999b]. The distributions of participating neurons, number of spikes per event, and fraction of events in which a single cell spikes are log-normal [Mizuseki and Buzsáki, 2013], strongly skewed with long tails. This means that there is a large variability in the recruitment of different units. Additionally, the skewed distribution of SWR intervals [Axmacher et al., 2008; Sullivan et al., 2011] suggests that SWRs resemble a random process.

In support of the hypothesis that SWRs originate in the hippocampus, multiple experiments have shown that SWRs persist to lesions of the neocortex [Buzsáki et al., 1983; Suzuki and Smith, 1988], and removal of the entorhinal cortex [Bragin et al., 1995b], the medial septum, and fimbra-fornix area [Buzsáki et al., 1983; Suzuki and Smith, 1988].

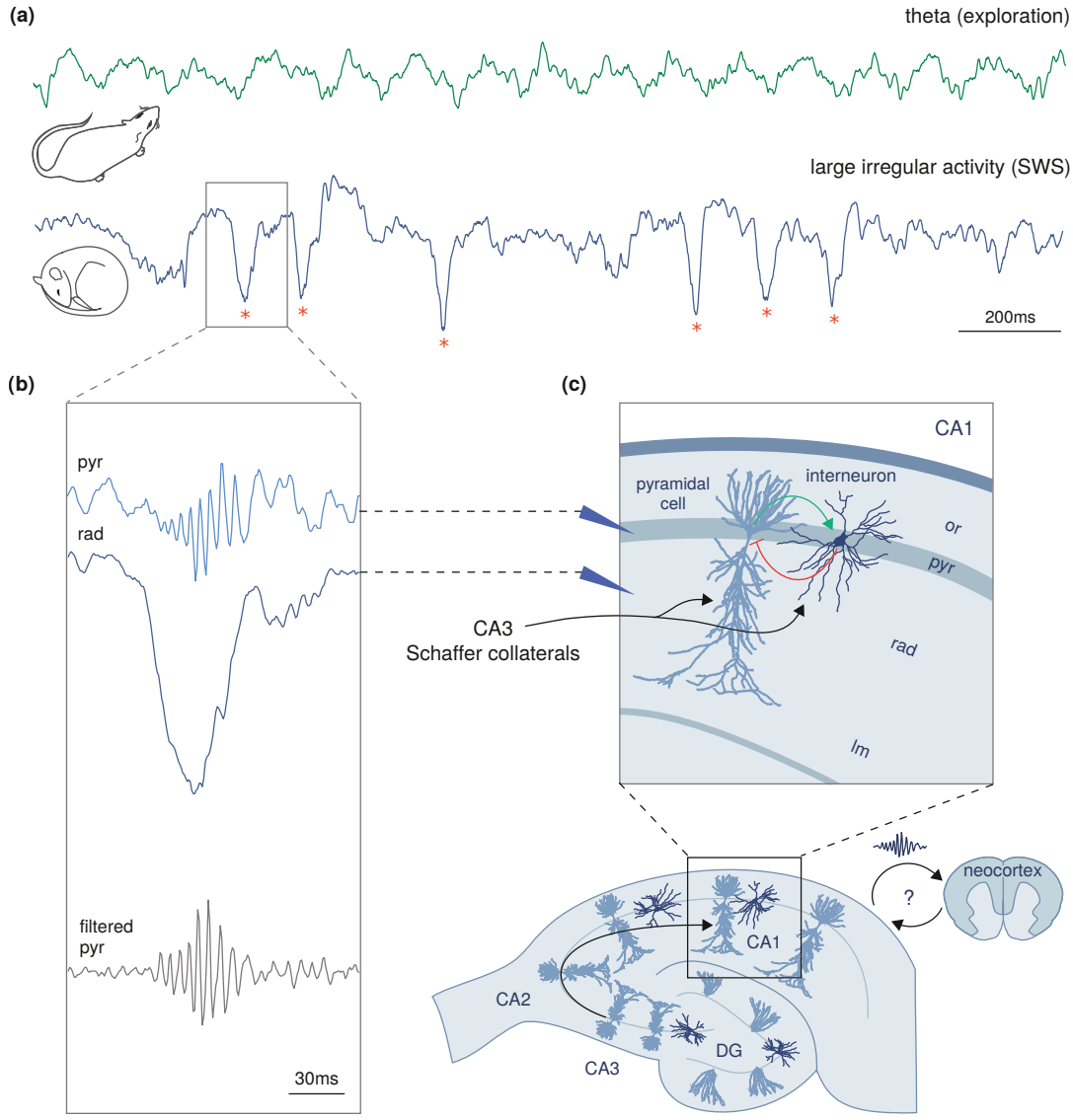

FIGURE 1.2: **SWRs and theta oscillations are visible in the hippocampal LFP during exploration and offline states. (a)** During exploration, the hippocampal LFP is characterized by theta oscillations (8-12 Hz). The activity becomes more irregular during periods of sleep and consummatory behavior. Red asterisks mark SWR events. **(b)** Depth profile of an isolated SWR event. In *stratum pyramidale* (pyr) both sharp wave and ripple components are visible. A large transient (the sharp wave) characterizes the signal in *stratum radiatum* (rad). In the lower plot, the signal is filtered in the 150-250 Hz frequency-band to isolate the ripple component. **(c)** Upper plot: schematic representation of CA1 hippocampal layers, cells arrangement and incoming CA3 input (Schaffer collaterals). The interaction between pyramidal cells and interneurons (red and green arrows) is believed to be important for the emergence of SWRs. Or, *stratum oriens*; pyr, *stratum pyramidale*; rad, *stratum radiatum*; lm, *stratum lacunosum moleculare*. Lower plot: schematic representation of a hippocampal slice depicting cells arrangement, directionality of propagation from CA3 to CA1 (Schaffer collaterals) and possible SWR-mediated interaction with neocortical areas. Adapted from Girardeau and Zugaro [2011] with permission from Elsevier.

Temporal analysis of SWRs in different hippocampal subfields and cut experiments revealed that SWRs originate in the CA3 area and successively spread to CA1 [Buzsáki, 1986; Sullivan et al., 2011; Davoudi and Foster, 2019]. As highlighted in Sec. 1.1.2, anatomical evidence support the directionality of propagation, even though this idea was challenged in a recent publication [Oliva et al., 2016]. Interestingly, CA1 can still generate sharp-wave-like events when CA3 inputs are blocked [Nakashiba et al., 2009]. The sharp wave component has been shown to be coherent across CA3 and CA1, while it has been hypothesized that the ripple component is regenerated in each region independently [Sullivan et al., 2011]. A recent study [Patel et al., 2013] addressed the question of ripple propagation along the septo-temporal axis and showed that multiple routes are possible, with ripples locally generated across the whole axis.

Overall, SWRs are a cooperative, emergent feature of hippocampal networks, and the fact that they can persist in the isolated hippocampus suggests that they form an intrinsic, default network state of the system.

1.1.3.3 The role of SWRs in memory consolidation

Even though the functional implications of SWRs are still unclear, evidence suggests that these rhythms could mediate the memory consolidation process. A groundbreaking study by Wilson and McNaughton [1994] showed that place cells are reactivated during periods of slow-wave sleep, and that this reactivation preferentially occurs during SWRs. In a follow-up study, Skaggs and McNaughton [1996] observed that the temporal order of cells' firing is preserved during exploration and slow-wave sleep (see Fig. 1.3). Since then, many studies have corroborated these findings (see e.g. [Kudrimoti et al., 1999; Nádasdy et al., 1999; Hirase et al., 2001; Lee and Wilson, 2002]) establishing the idea of time-compressed sequence replay during SWRs. In this context, the fast ripple oscillations could mediate the time compression of sequences. Additionally, O'Neill et al. [2008]; Cheng and Frank [2008]; Giri et al. [2018]; Hwaun and Colgin [2019] showed that SWR-associated replay occurs more frequently after an animal explored a novel environment, and observed that place cells encoding novel environments are more active during SWRs than those encoding familiar environments, hinting at a consolidation process of recently experienced memories. These results support the two-stage model hypothesis, but more work needs to be done to address the question of reactivation of more complex memories, and memories that involve non-spatial components. Some steps in this direction were done by Dave and Margoliash [2000]; Hahnloser et al. [2002] in songbirds, Hoffman and McNaughton [2002] in primates, and Staresina et al. [2013] in humans.

Is SWR occurrence linked to memory performance? Several experiments have addressed this question and have shown that, in rodents, the number of SWRs increases during post-learning sleep [Eschenko et al., 2008; Girardeau and Cei, 2014], and that it can be used to predict memory performance [Dupret et al., 2010]. In humans, Axmacher et al. [2008] found a positive correlation between the number of successfully learned items

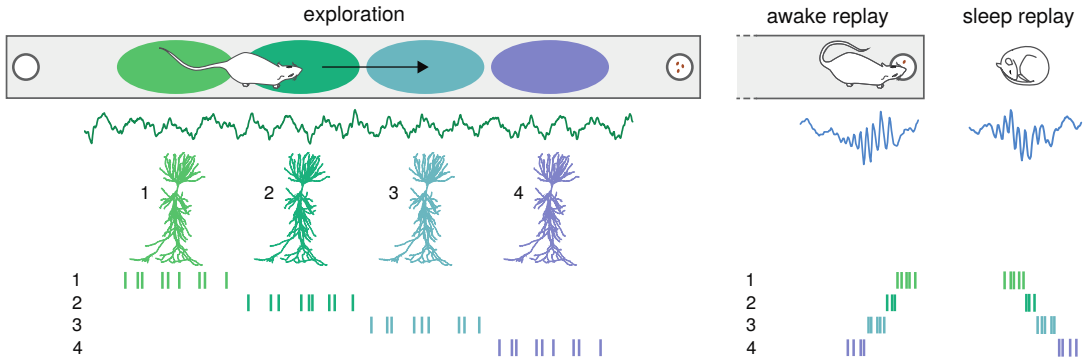

FIGURE 1.3: **Place cell sequences are reactivated during offline states**. Left: schematic representation of a rat running on a linear track during the exploration phase. Four place cells are represented, that preferentially fire in the color-coded place fields on the track. As the animal runs to the reward location, the four cells are active sequentially. This happens during theta oscillations (green trace). Right: during SWR episodes happening in offline states of awake rest and sleep, the place cell sequences are reactivated in a time-compressed manner, in reverse and forward order, respectively. Adapted from Girardeau and Zugaro [2011] with permission from Elsevier.

and the density of post-learning ripples in the rhinal cortex (but, interestingly, not in the hippocampus).

To tighten the causal link between SWR-mediated reactivation and memory performance, Girardeau et al. [2009] tested the effect of post-learning blockage of SWRs in rats. In animals trained on a behavioral task, SWRs were detected online during the first hour of post-training sleep and selectively blocked by electrical stimulation. The performance of these animals was compared to the performance of animals that were trained on the same task, but that received no electrical stimulation, or electrical stimulation following SWRs. The results presented in Fig. 1.4 show that SWR blockage impairs memory performance. A similar experiment [Ego-Stengel and Wilson, 2010] confirmed this finding.

In addition to SWRs occurring during sleep, attention has focused in recent years on awake SWRs. During awake SWRs, reverse-order reactivation (see Fig. 1.3) of past and current place fields has been reported [Foster and Wilson, 2006; Jackson et al., 2006; Diba and Buzsáki, 2007; Csicsvari et al., 2007; Karlsson and Frank, 2009; Davidson et al., 2009; Gupta et al., 2010], which is possibly linked not only to the consolidation of experienced trajectories, but also to the planning of new ones (see also Pfeiffer and Foster [2013]; Dragoi and Tonegawa [2011]). One important experiment in this direction was performed by Jadhav et al. [2012], who used a similar stimulation paradigm as the one described by Girardeau et al. [2009] and showed that disrupting awake ripples impairs spatial working memory, i.e., a memory-guided decision-making process, but not the memory of trajectories that do not involve planning. More recently, Wang et al. [2015] showed that suppressing awake SWRs by optogenetically activating the median raphe neurons prevents the consolidation of fear conditioning memories. Interestingly, Roux et al. [2017] showed that blocking awake SWRs by suppressing the firing of a small group

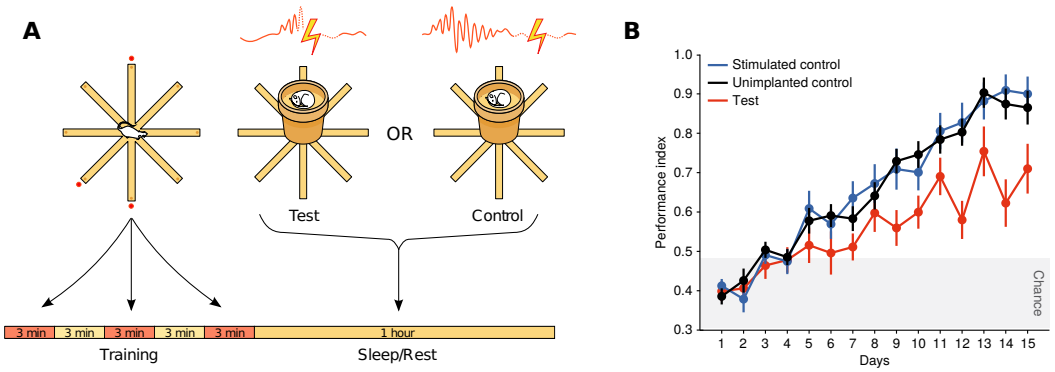

FIGURE 1.4: **Selective suppression of SWRs impairs memory consolidation. A)** Rats are trained to find rewards in an eight-arm radial maze. After the third trial, rats are allowed to sleep. Implanted animals receive electrical stimulations during the first hour of sleep, selectively during SWRs (test group) or following SWR occurrence (control). **B)** Memory performance is tested on multiple days and compared across implanted animals (test and control conditions), and unimplanted animals that received no stimulation. Memory performance is reduced for the test group (red). Adapted from Girardeau et al. [2009] with permission from Springer Nature.

of pyramidal cells impairs the stabilization of place fields representation without affecting memory performance.

Another feature that hints at an important role of SWRs in memory consolidation is the link with synaptic plasticity. As already mentioned, the presence of synaptic plasticity throughout the hippocampus supports the two-stage model hypothesis of the hippocampus as the place where memories are first formed. However, this does not necessarily imply that plasticity takes place, during the consolidation phase, in correspondence with SWRs. Relevant experiments in this direction were performed by Abraham et al. [1986]; Buzsáki et al. [1987], who showed that hippocampal population bursts (like SWRs) *in vitro* can induce LTP. Also the reverse seems to hold: Behrens et al. [2005] showed in hippocampal slices that NMDA-receptor-dependent induction of long-term potentiation in CA3 can induce SWRs. More importantly, Girardeau and Cei [2014] showed that post-learning increase in SWR incidence was impaired by blockade of NMDA-receptors during learning (but not after). NMDA receptors are known to be crucial for hippocampal synaptic potentiation [Morris et al., 1986] and have been shown to be linked to the successful consolidation of spatial memory [Tsien et al., 1996; Shimizu et al., 2000; Dupret et al., 2010] (see Nakazawa et al. [2004] for a review). Even though these results are not conclusive, they support the hypothesis that during consolidation, the reactivation of already created memories induces synaptic plasticity between connected neurons in a cell assembly.

1.1.4 Cortical rhythms

1.1.4.1 Gamma oscillations during exploratory behavior

In neocortical structures, the awake state is characterized by gamma oscillations (30-100 Hz) [Adrian, 1942; Bressler and Freeman, 2003; Gray and Singer, 1989; Maloney et al., 1997]. Compared to hippocampal oscillations, cortical gamma oscillations have been shown to be more transient and localized, both in rodents and human studies [Edwards et al., 2005; Canolty et al., 2006; Sirota et al., 2008]. Even though gamma oscillations have been observed in various cortical regions (visual, auditory, somatosensory and motor cortices just to name a few), their functional role remains to date obscure (see Merker [2013] for a review). At the cellular level, neurons have been reported to be in a desynchronized state [Harris and Thiele, 2011; Poulet and Crochet, 2019], with little shared subthreshold fluctuations in their membrane potentials. Interestingly, Sirota et al. [2008] showed that cortical gamma oscillations, and cortical unit firing, can be entrained by hippocampal theta oscillations. Similar results were reported by Canolty et al. [2006] in humans.

1.1.4.2 Slow oscillations and thalamic spindles during offline states

During slow-wave sleep, the LFP activity of neocortical areas and EC is characterized by slow oscillations (0.5-2 Hz) [Steriade et al., 1993b,d,c; Achermann and Borbély, 1997]. At the cellular level, principal cells alternate between depolarized (UP) and hyperpolarized (DOWN) states (see Fig. 1.5A). The alternation between the two states is synchronous across cortical areas [Volgushev et al., 2006] and it is thought to reflect a bistable mechanism, driven by either cell intrinsic or network-wide mechanisms [Compte et al., 2003; Holcman and Tsodyks, 2006; Jercog et al., 2017; Levenstein et al., 2019].

A large body of research has focused on understanding how slow oscillations are linked to the occurrence of SWRs, and how the coordination of these rhythms may affect memory consolidation. Sirota et al. [2003]; Isomura et al. [2006]; Mölle et al. [2006]; Clemens et al. [2011] have shown that the excitability changes in the cortex can affect SWR occurrence. Many studies have looked at the temporal relations between SWRs and UP-DOWN states, but a strong consensus is still missing. Battaglia et al. [2004]; Isomura et al. [2006]; Jiang et al. [2017] showed that SWRs preferentially occur at the DOWN-to-UP transition, whereas Sirota et al. [2003]; Peyrache et al. [2011] suggested that SWRs are more likely to occur at the end of an UP state. Additionally, Sirota et al. [2003]; Isomura et al. [2006] showed that the co-occurrence of SWRs with UP states in EC is stronger than with UP states in the neocortex, suggesting that slow oscillations might spread from the neocortex to the EC and from there pace hippocampal SWRs. If this is the case, the Poisson-like distribution of UP and DOWN states transitions could explain the irregular occurrence of SWRs [Buzsáki, 1986].

At the same time, evidence suggests that the SWR output could influence cortical activity. Chrobak and Buzsáki [1994] showed that SWRs can affect the firing of cells in

the deep layers of EC during both UP and DOWN states, even though the effect seems to be insufficient to drive DOWN-UP transitions [Isomura et al., 2006]. It remains to be clarified how much these results are affected by the anesthetized state of the animals.

Additional to slow oscillations, the early (superficial) stages of sleep are characterized by the presence of sleep spindles, oscillations at 12-18 Hz that last 1-3 seconds and are known to originate from the thalamus [Steriade et al., 1993b] during UP states. More specifically, they originate from the interaction between GABAergic neurons of the thalamic reticular nucleus and glutamatergic thalamic cells [Steriade et al., 1993d,c; De Gennaro and Ferrara, 2003; Timofeev and Bazhenov, 2005; Steriade, 2006] and are known to reach the hippocampus either via the EC or by a direct thalamic connections to CA1. It has been shown [Sirota et al., 2003; Sullivan et al., 2014] that thalamocortical spindles modulate the firing rate of hippocampal cells and EC cells in layers II, III, and V. Vice versa, SWRs increase neuronal firing and spindle power in prefrontal cortex [Wierzynski et al., 2009]. The exact temporal relationship between spindles and SWRs remains unclear: SWRs precede spindles on a time scale of seconds [Siapas and Wilson, 1998], but, at a finer (millisecond) timescale, SWRs are often phase-locked to spindle cycles [Sirota et al., 2003; Mölle et al., 2006; Clemens et al., 2007], suggesting that spindles might influence SWR timing. One confounding factor might be the cortical region of interest: spindles recorded in the visual and somatosensory cortices are more likely to entrain SWRs [Sirota et al., 2003; Ji and Wilson, 2007], whereas prefrontal spindles follow SWRs [Peyrache et al., 2011].

Finally, K-complexes are sharp-wave-like events happening at the DOWN-UP transition of slow oscillations [Amzica and Steriade, 1997; Steriade and Amzica, 1998] that, by synchronizing the network and activating the GABAergic neurons of the thalamic reticular nucleus, are thought to induce sleep spindles [Steriade et al., 1993a].

Even though many details are yet to be clarified, the existence of a bidirectional influence between hippocampal and cortical oscillations supports the hypothesis of a transfer of information between the two areas: potentially, cortical oscillations could recruit hippocampal neurons to start SWR events, and the information replayed during SWR events could be transferred back to cortical regions. An intriguing hypothesis is that, due to the transient nature of SWRs (compared to the more long-lasting cortical events), a SWR could influence the same cortical oscillation that favored its initiation [Sirota and Buzsáki, 2005; Mölle et al., 2009; Sullivan et al., 2011; Buzsáki, 2015].

1.1.4.3 Cortico-hippocampal loop and memory consolidation

To test the hypothesis of a SWR-mediated transfer of information from the hippocampus to the neocortex, Ji and Wilson [2007] simultaneously recorded hippocampal and visual cortical cells and showed that neuronal ensembles are reactivated in parallel during slow-wave sleep (see Fig. 1.5B, C). These results on non-hippocampal replay were corroborated by studies in prefrontal cortex and subcortical structures like the ventral striatum and forebrain [Ribeiro et al., 2004; Euston et al., 2007; Lansink et al., 2008]. Recently,

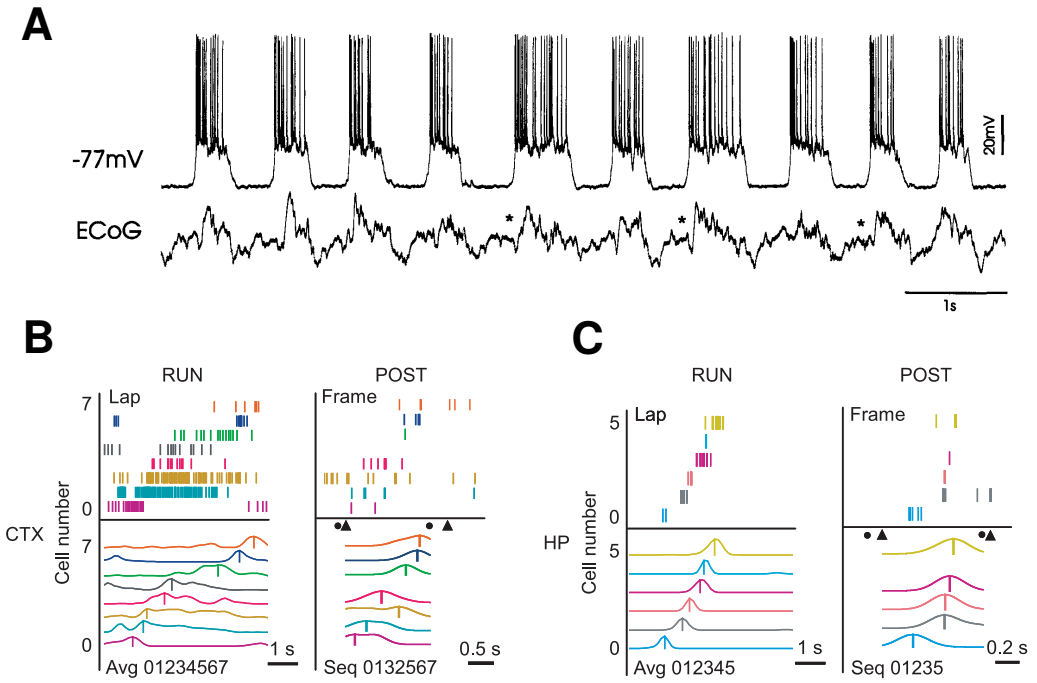

FIGURE 1.5: **Cortical dynamics during offline states. A)** Top: *in vivo* intracellular recording of a pyramidal cell in the prefrontal cortex of a rat (anesthetized rat, recording at resting membrane potential, −77 mV) shows spontaneous switches between epochs of low activity (DOWN states) and epochs of depolarized state with intense firing (UP states). Bottom: slow oscillations visible in the electrocorticography (ECoG, or intracranial EEG) of the contralateral cortex. Asterisks mark locking between positive deflection in the ECoG and DOWN-to-UP transitions. Adapted from Branchereau et al. [1996] with permission from John Wiley and Sons. **B) - C)** Coincident replay in hippocampus and visual cortex. B) Cortical firing sequences during RUN and POST sleep epoch. Left: the upper plot shows the spiking activity of the cells over a single lap, the lower plot shows the average firing sequence across all laps. There is a clear sequence emerging in the RUN (cells are numbered according to their peak of firing). Right: upper plot shows the spiking activity of the same cells shown on the left during a POST sleep frame. Triangles and circles depict the frame's start and end times. The lower plot shows the firing sequence in the frame (spikes are convolved with a Gaussian filter). Note that the RUN sequence is partially replayed during the POST sleep epoch. C): same as B), for cells in the hippocampus coding for the same trajectory. Adapted from Ji and Wilson [2007] with permission from Springer Nature.

Jiang et al. [2017] showed replay of brain-wide sequences of activation that preferentially occurs during SWRs in humans. Even if it might be tempting to conclude that the hippocampus is driving cortical and subcortical reactivation, it is probably not the only drive, as observed in post-learning replay in the entorhinal cortex by O'Neill et al. [2017].

Do cortical oscillations influence memory consolidation? To address this question, Marshall et al. [2006] used transcranial brain stimulation in humans to boost slow oscillations and showed that this manipulation improved the retention of declarative memories (the result was dependent on the stimulation frequency). Comparable results were obtained by Ngo et al. [2013] using auditory stimulation to enhance slow oscillations, and by Binder et al. [2012, 2014] in rats.

To investigate the joint contribution of SWRs, slow oscillations, and thalamocortical spindles in memory consolidation, Latchoumane et al. [2017] studied the effect of inducing spindles on a hippocampus-dependent fear-conditioning task. This study demonstrated that only spindles induced during slow oscillations' UP states (in-phase spindles) enhance memory performance (stimulating during DOWN states had no effect), and that memory impairment can be induced by blocking the generation of spindles during the UP state. Induced in-phase spindles enhanced the co-occurrence of ripples and slow oscillations, suggesting that the concerted activation of the three rhythms is a prerequisite for the successful induction of memory consolidation.

1.2 The question of SWR generation

In this work, I address the question of how SWRs can be generated in a computational model. The evidence suggesting that SWRs are an intrinsic phenomenon of the hippocampus (see Sec. 1.1.3.2) motivates the choice of modeling this area and neglecting external (cortical) inputs. In Section 1.2.1, I review experimental evidence showing that SWRs emerge spontaneously in *in vitro* preparations. Afterward I discuss existing hypotheses on SWR generation (Sec. 1.2.2), and show that the *in vitro* setup provides insights on the circuit mechanisms supporting SWRs (Sec. 1.2.3). Finally, Section 1.2.4 discusses computational models of SWR generation and the main SWR features not yet explained by the existing literature.

1.2.1 Using an *in vitro* preparation to study SWRs

To address the question of how SWRs can be generated in the hippocampus, we need to understand which is the microcircuit that supports these events: what are the cells present in the network, how are they connected to each other, which is the effect of specific synaptic manipulations onto the existence of SWRs, etc. Questions like these are very difficult to address in a behaving animal, because we only have limited access to the network and are mostly agnostic to which type of cells we are recording from. Moreover, it is difficult to study the network connectivity in an *in vivo* setup. Luckily, many of these questions can be answered using an *in vitro* slice preparation.

SWRs have been shown to emerge spontaneously in hippocampal slices [Maier et al., 2002, 2003; Hájos et al., 2009; Ellender et al., 2010] (see Fig. 1.6A-D). Typically, experimentalists use horizontal slices from the ventral hippocampus (where cell excitability has been shown to be larger [Dougherty et al., 2012]), 300-500 μm thick. *In vitro* SWRs share many features with *in vivo* SWRs. Their incidence is similar: 0.01-3/s *in vivo*, [Buzsáki, 1986], and 2.8 ± 0.2/s and 2.7 ± 1.1/s *in vitro* (mean $\pm$ SEM), Papatheodoropoulos and Kostopoulos [2002] and Maier et al. [2003] respectively (statistics from Maier and Kempter [2017]). The oscillatory frequency of ripples is 150-250 Hz *in vivo* [Buzsáki et al., 1992; Ponomarenko et al., 2004; Sullivan et al., 2011] and 210 ± 16 Hz *in vitro* [Maier et al., 2003] (but see Maier et al. [2011] for evidence of lower ripple frequency *in vitro*).

Additionally, the laminar profile of SWRs is preserved *in vitro* [Maier et al., 2009; Schönberger et al., 2014; Hofer et al., 2015], even though the SWR amplitude may vary (see Fig. 1.6F-I). As *in vivo*, SWR events *in vitro* have been shown to originate in the CA3 area and to then propagate to the CA1 and downstream areas [Wu et al., 2002; Maier et al., 2003; Kubota et al., 2003; Both et al., 2008; Maier et al., 2009; Ellender et al., 2010]. Still, CA1 seems capable of sporadically generating SWR events in absence of CA3 inputs [Maier et al., 2003, 2011] (but see Colgin et al. [2004]; Ellender et al. [2010]).

One potential difference between the *in vivo* and *in vitro* scenarios is that *in vivo* SWRs are characterized by a strong irregularity [Buzsáki, 1986], which has been reported only partially *in vitro* [Papatheodoropoulos, 2010; Schlingloff et al., 2014] (but high regularity in Behrens et al. [2007]). This difference could result from the lack of neocortical inputs in the slice setup.

Even though the experiments presented here point at the *in vitro* model as a scenario where physiological (*in vivo*-like) SWRs can be observed, it is important to keep in mind that the existence of spontaneous SWRs is influenced by different conditions, as animal species, slicing angle, slice thickness, composition of the extracellular fluid, and slice position with respect to the dorso-ventral hippocampal axis. Many of these confounding factors are discussed by Buzsáki [2015].

The existence of intrinsically generated SWRs in the slice preparation facilitates the study of the mechanisms regulating their occurrence. In Sections 1.2.2 and 1.2.3, I show that even though *in vitro* results cannot explain all aspects of SWR generation, they provide a useful tool to better understand the neuronal circuitry involved.

1.2.2 Hypotheses on SWR generation

Given the ubiquitous presence of SWRs *in vivo* and *in vitro*, across multiple hippocampal regions and recording conditions, there has been an increasing interest in understanding the mechanisms underlying the emergence of SWRs.

From the perspective of the two-stage model, the main questions relate to the alternation of theta and non-theta periods: how does the behavior (exploration vs. rest) affect the occurrence of different brain signals (theta-gamma vs. SWRs)? Michael Hasselmo was one of the first to propose that neuromodulation could play an important role. More specifically, Hasselmo [1999] proposed that acetylcholine (ACh), a neurotransmitter whose levels have been shown to vary between awake states, slow-wave sleep and REM sleep, could mediate the transition from theta to SWR states. During awake states, ACh levels in hippocampus and cortex have been shown to be high, and be drastically reduced during slow-wave sleep [Kametani and Kawamura, 1990; Marrosu et al., 1995]. The presence of ACh seems to favor the emergence of theta oscillations [Monmaur et al., 1996], and ACh has been shown to suppress glutamatergic synaptic transmission inside the hippocampus and from the hippocampus to the EC [Valentino

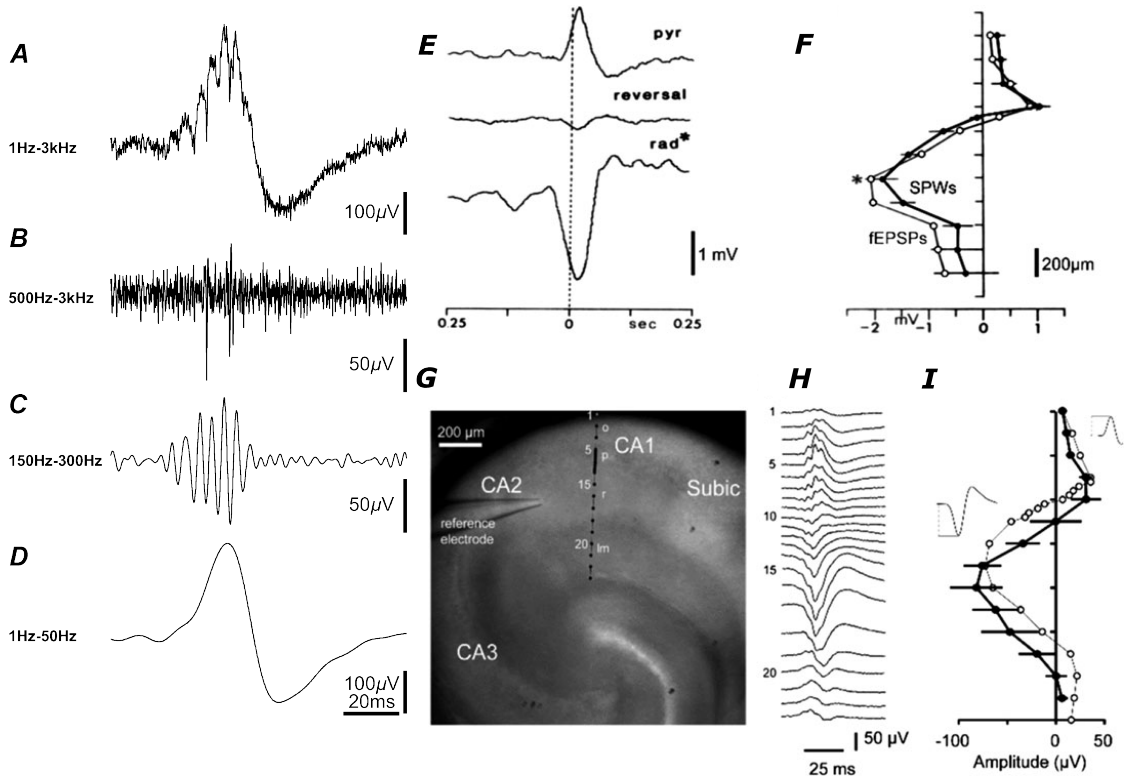

FIGURE 1.6: **SWRs *in vitro* are similar to *in vivo* SWRs. A) - D)** An isolated SWR event recorded *in vitro* in CA1 *stratum pyramidale*. Band-pass filtering in different frequency bands (frequency range specified to the left) reveals the ripple (C) and sharp wave (D) components. Occasionally, units can be visible when filtering at high frequencies (B). Reproduced from Maier et al. [2003] with permission from John Wiley and Sons. **E) - F)** Depth profile of SWRs *in vivo*. E) Averaged (across 128 sweeps) and band-pass filtered (1-30 Hz) CA1 SWRs *in vivo* show phase reversal in *stratum pyramidale* (pyr) and *stratum radiatum* (rad). Events are aligned with respect to multi-unit activity recorded by a reference electrode in *stratum pyramidale*. F) Laminar profile of SWR amplitude (SWs, black filled circles) and EPSPs evoked by Schaffer collaterals stimulation (fEPSPs, white open circles). Ticks on the vertical axis are 166 μm apart. **G) - I)** Depth profile of SWRs *in vitro* shares features with *in vivo* recordings. G) Infrared image showing a ventral hippocampal slice, reference electrode at the CA2-CA1 border, and recording electrode in CA1 with multiple recording sites (black dots are 100 μm apart). Strata abbreviations: o, *oriens*; p, *pyramidale*; r, *radiatum*; lm, *lacunosum moleculare*. H) LFP averages (20 events) with recording electrode presented in G. Events are aligned with respect to the signal at the reference electrode. I) SWR amplitude profile of the recording shown in H) (open circles), and an average over 5 recordings (black filled circles). Insets show example traces of the SWR positive and negative peaks. Figure reproduced from Maier and Kempter [2017] with permission from Springer Nature. (E)-F), original image from Buzsáki et al. [1983]; G)-I), original image from Maier et al. [2009]).

and Dingledine, 2018; Sheridan and Sutor, 1990; Hasselmo, 1995], while only mildly affecting transmission from the superficial layers of the EC to the hippocampus [Hasselmo and Schnell, 1994; Hasselmo et al., 1995]. These studies suggest that ACh could gate the transmission of information from the EC to the hippocampus to allow for the encoding of new memories, while blocking internal hippocampal inputs and inputs from the hippocampus to cortical areas, which could interfere with the encoding process. The drop of ACh levels during slow-wave sleep could allow for the intrinsic generation of SWRs in the hippocampal network, a hypothesis that is supported by experiments showing that the activation of cholinergic neurons (i.e., neurons releasing ACh) in the medial septum strongly suppresses SWRs [Vandecasteele et al., 2014; Hunt et al., 2018; Zhou et al., 2019]. Possibly, low ACh levels could favor the communication from the hippocampus to the EC, strengthening the hypothesis of a transfer of memory traces across the two areas. Despite these promising results, direct evidence of a link between ACh modulation and memory performance is still lacking.

Ever since their discovery, it has been proposed that SWRs are generated by a build-up of activity in the extensive CA3 recurrent system [Buzsáki et al., 1983; Menendez de La Prida et al., 2006], with activity propagating to the CA1 area. In support of this hypothesis, CA3 was believed to be strongly recurrently connected [Miles and Wong, 1986; Amaral and Witter, 1989; Ishizuka et al., 1990; Witter, 2007], but recent evidence has shown that the recurrent connectivity could be as low as ∼1% [Guzman et al., 2016].

In parallel, other studies hypothesized an important role for interneurons in the initial phase of the SWR [Ellender et al., 2010; Sasaki et al., 2014; Schlingloff et al., 2014; Bazelot et al., 2016]. How can the early involvement of interneurons be linked to the initiation of a population burst? To answer this question, we first need to understand how interneurons are embedded in the hippocampal microcircuit responsible for SWR generation, an aspect that I address in the next section.

1.2.3 Understanding the microcircuit

As discussed in Section 1.2.1, the *in vitro* model is a useful tool to study the neuronal mechanisms supporting SWR generation. Hence, this section presents (mostly) *in vitro* experimental results that are relevant to model SWRs. In particular, I focus on the cell types whose firing has been shown to be modulated by SWRs, and on the main factors that are thought to regulate SWR occurrence. I am interested in defining a model that exhibits spontaneous SWRs with realistic dynamics, as well as reproducing existing evidence suggesting that SWRs can be triggered by cell stimulation.

Differential firing of cells in SWR and non-SWR periods. SWRs events are characterized by synchronous activation of ∼10% of hippocampal neurons [Csicsvari et al., 2000] (with up to ∼50% of units involved [Ylinen et al., 1995; Mizuseki and Buzsáki, 2013]). Different cell types are differently modulated by SWR events (see Fig. 2.3). Pyramidal cells in CA1 and CA3 fire sparsely when no SWR occurs, and increase their

firing (∼6-fold) in correspondence with an event [Csicsvari et al., 1999b; Stark et al., 2014]. Mean firing rates for non-SWR periods are in the range of 0.03-3 spikes/s, and in the range of 1-13 spikes/s for SWR period, in both CA1 [Ylinen et al., 1995; Csicsvari et al., 2000; Klausberger et al., 2003; Lapray et al., 2012; English et al., 2014] and CA3 [Ellender et al., 2010; Lasztoczi et al., 2011; Hájos et al., 2013; Tukker et al., 2013; Csicsvari et al., 2000]. Pyramidal cells mostly fire once during a SWR event [Ylinen et al., 1995; Hájos et al., 2013]; *in vivo*, however, the firing rate can reach 40 spikes per second [English et al., 2014; Hulse et al., 2016]. The participation of pyramidal cells to SWRs in humans seems to be comparable to the results in rodents [Le Van Quyen et al., 2008].

Similar to pyramidal cells, parvalbumin-positive basket cells (PV^+BCs) have been shown to increase their firing activity (∼3-fold) during SWRs [Csicsvari et al., 1999b], while being almost silent in non-SWR periods [Csicsvari et al., 1999b; Klausberger and Somogyi, 2008]. Mean firing rates are in the range of 3-20 spikes/s in non-SWR periods and increase up to 120 spikes/s during SWRs, in CA1 [Klausberger et al., 2003; Lapray et al., 2012; Varga et al., 2012; Royer et al., 2012; Varga et al., 2014; Stark et al., 2014] and CA3 [Hájos et al., 2013; Tukker et al., 2013; Schlingloff et al., 2014].

Moreover, somatostatin-positive (SOM^+) cells like bistratified and oriens-lacunosum-moleculare (O-LM) interneurons have been shown to increase their firing at different phases of the SWR events [Varga et al., 2012; Pangalos et al., 2013; Varga et al., 2014; Katona et al., 2014, 2017] (even though O-LM were reported to be suppressed in Klausberger et al. [2003]). Other types of interneurons, like cholecystokinin-positive (CCK^+) basket cells and ivy cells, do not seem to be strongly modulated by SWRs [Klausberger et al., 2005; Lasztoczi et al., 2011].

The modulation of axo-axonic cells (AACs) by SWRs is more controversial. Klausberger et al. [2003]; Varga et al. [2012]; Hájos et al. [2013]; Varga et al. [2014] showed that AACs increase their firing probability right before, or at the beginning, of a SWR event, but are silent during the later part of, and shortly after, the event. At the same time, the firing of AACs seems to be suppressed during SWRs in Viney et al. [2013], with no increase in firing preceding an event and irregular epochs of high firing in non-SWR periods (up to 30 spikes/s). Finally, Fuentealba et al. [2008] identified an enkephalin-expressing (ENK) GABAergic cell in CA1 that is active during non-SWR periods (with a strong post-SWR activity rebound) and stops firing during the event. More data are needed to better clarify the role of ENK cells during SWRs. In line with this work, Csicsvari et al. [1999b] reported the existence in CA1 of anti-SWR interneurons. However, the lack of details about the interneurons' identity prevents a clear link to the groups discussed here. Similarly, Le Van Quyen et al. [2008] showed that a subclass of putative interneurons recorded in the human hippocampal formation stopped firing at the beginning of a SWR event.

Lastly, an interesting property of cells' firing during SWRs is their ripple modulation. In fact, the majority of both pyramidal cells and interneurons have been shown to fire locked to the ripple cycle [Buzsáki et al., 1992; Csicsvari et al., 1999a; Le Van Quyen

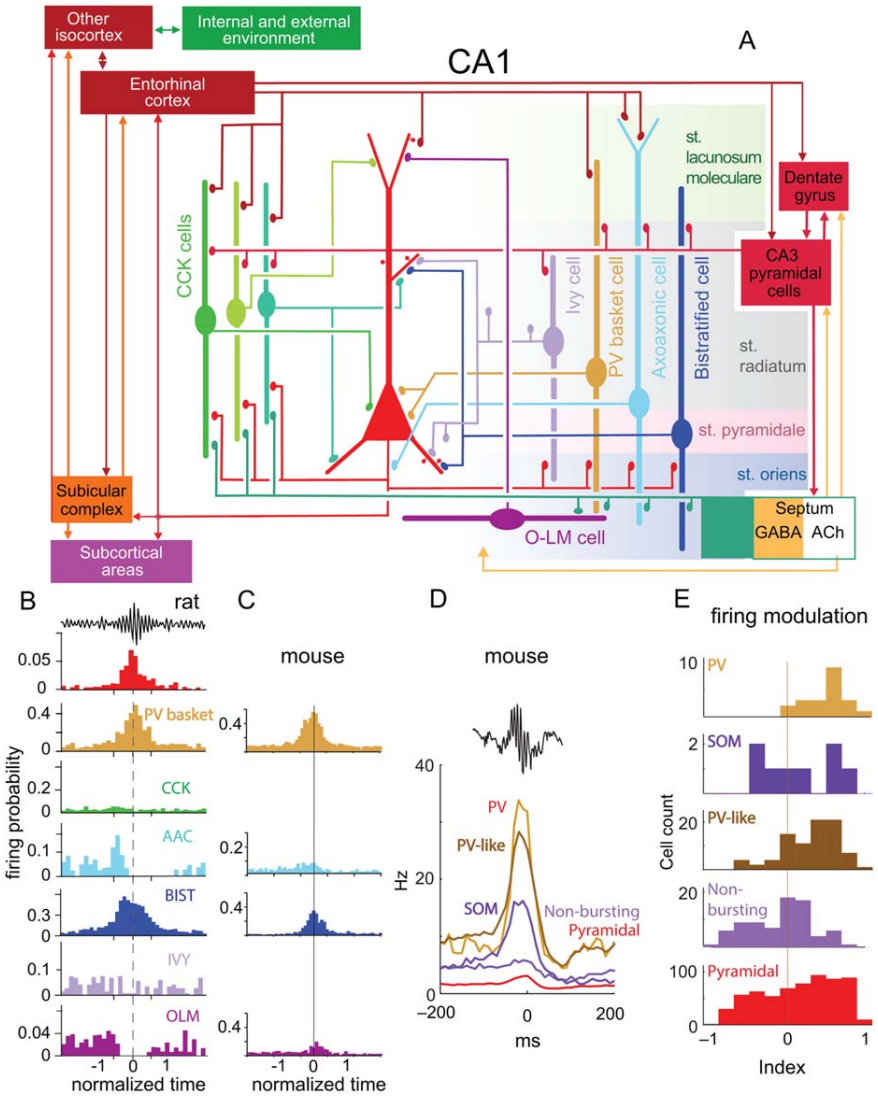

FIGURE 1.7: **Interneurons modulation during SWR events. A)** Schematic representation of connections between a CA1 pyramidal cell (red cell with triangular cell body) and interneurons, and main input and output regions of CA1. Connections among interneurons are not shown. **B)** Firing probability histograms of different cell types, aligned to the ripple peak. Data were recorded in anesthetized rats. **C)** Same as B), but for data recorded in non-anesthetized mice. **D)** Firing rates of optogenetically identified PV^+ and SOM^+ cells in behaving mice, aligned to ripple peak. **E)** Firing modulation of cells with respect to ripples. Positive indices indicate an increase of firing during ripples, negative indices indicate firing reduction during ripples. Reproduced from Buzsáki [2015]. A) and B), original image from Somogyi et al. [2014]; C) original image from Varga et al. [2012]; D) and E), original image from Royer et al. [2012].

et al., 2008; Bähner et al., 2011; Hájos et al., 2013; Stark et al., 2014], suggesting that not only the firing rate, but also the timing of spiking, might influence the contribution of different cell types to the consolidation process.

SWRs can be triggered by cell stimulation. Even though SWR events happen spontaneously in a slice, several experiments have shown that they can be triggered by stimulating different cell types. Indeed, Bazelot et al. [2016] showed that activating even

a single CA3 pyramidal cell might be sufficient to initiate a SWR event. By driving a pyramidal neuron to spike a single action potential (by injecting a depolarizing current), the authors showed that a SWR event could be elicited with latencies $\leq$ 5 ms. Interestingly, the stimulation of pyramidal cells triggered the activation of several perisomatic interneurons, which suggests that the recruitment of interneurons is important for SWR generation. Similarly, but *in vivo*, Stark et al. [2014] showed that local optogenetic activation of CA1 pyramidal cells results in the generation of SWR-like events (termed high-frequency oscillations). Additionally, weak electrical stimulation in CA3 *stratum pyramidale* has been shown to elicit SWR events *in vitro* with latencies ∼15-80 ms, depending on the stimulus strength [Jiang et al., 2018]. Note that, in this scenario, multiple cell types (i.e., not only pyramidal cells) might have been stimulated.

What is the role of interneurons in SWR generation? To answer this question, Schlingloff et al. [2014] focused on CA3 PV^+ interneurons, that have been shown to be the most active class of interneuron during SWRs. A brief whole-slice light stimulation to PV^+ cells *in vitro* triggered LFP events identical to spontaneous SWRs, regardless of stimulation length (5–50 ms, see Fig. 1.8B-D). This result, replicated by Kohus et al. [2016], is counterintuitive because the activation of PV^+ interneurons results in a network state (a SWR) where pyramidal cells increase their firing (as described above). Interestingly, Schlingloff et al. [2014] showed that shortly ($<$ 10 ms) after stimulation, PV^+ cells receive excitatory postsynaptic currents (EPSCs), confirming the recruitment of excitatory cells. Moreover, optogenetic silencing of PV^+ cells paired with online SWR detection showed that silencing PV^+ cells interrupts SWR events. Additionally, the likelihood of observing spontaneous SWRs is strongly decreased during periods of PV^+ cells silencing (see Fig. 1.8F-G). Even though the mechanisms remain elusive, this set of experiments suggests that PV^+ cells play a crucial role in the generation of SWRs. Importantly, [Stark et al., 2014] failed to replicate the experiments relative to PV^+ activation in CA1, *in vivo*. One possible explanation could be that, due to the locality of optogenetic stimulation, not enough PV^+ cells were recruited to start an event in Stark et al. [2014] (it is estimated that approximately 20 interneurons were activated by light stimulation). For a detailed modeling study on the estimated number of PV^+ cells needed to trigger a ripple event, see Holzbecher and Kempter [2018].

What regulates the occurrence of SWR events? A puzzling feature of spontaneously occurring SWRs is the coexistence of two different timescales: while a single event lasts 50-100 ms, the incidence of events is in the range of 1-2/s (depending on preparation and recording site [Buzsáki, 2015]). Short-term synaptic depression has been suggested as a mechanism that could help bridging the gap between the two timescales. In a recent paper, Kohus et al. [2016] recorded spontaneous events *in vitro* in CA3 and observed a strong correlation between the amplitude of SWR events and the interval to the previous SWR (inter-event-interval, or IEI), recorded in the LFP (see Fig. 1.9A and Jiang et al. [2018] for similar results in CA1). Interestingly, the authors suggested that the existence

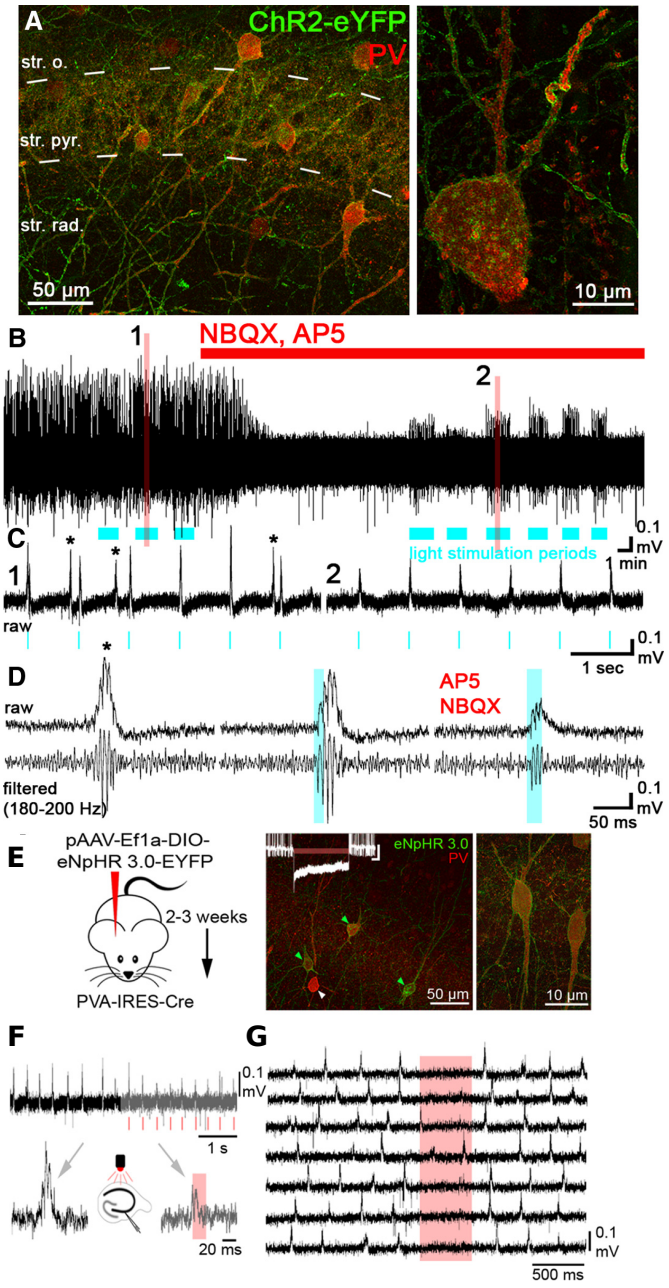

FIGURE 1.8: **Activation of PV$^+$ cells triggers SWRs and their blockage interrupts and prevents SWRs. A)** Double staining for PV (red) and ChR2 (green) shows that only PV$^+$ cells express ChR2 in this set of experiments. Str. o., *stratum oriens*; str. pyr., *stratum pyramidale*; str. rad., *stratum radiatum*. **B)** LFP shows spontaneous and evoked (by optogenetic activation of PV$^+$ cells, blue bars) SWR events, in control condition (left) and when excitatory synaptic transmission in blocked (NBQX, AP5 condition, red line). **C)** Two enlarged views of the LFP are presented, corresponding to periods denoted with 1 and 2 in B). Events marked with a star represent spontaneous events. Note that when excitatory transmission is blocked (right plot) spontaneous events disappear. **D)** SWR events are shown in isolation in the three conditions: spontaneous, evoked, and evoked with excitatory transmission blocked. Note the similarity between spontaneous and evoked events. **E)** Left: immunofluorescent construct is injected in a mouse Cre line to selectively silence PV$^+$ cells in the CA3 area. Right: immunofluorescent staining reveals double expression in PV$^+$ cells. Inset: example of a PV$^+$ neuron effectively silenced by illumination (current-clamp configuration). **F)** Online detection of spontaneously occurring SWRs triggers a transient laser illumination (red) that interrupts ongoing SWRs. Left: spontaneous SWR. Right: SWR interrupted by PV$^+$ cells silencing driven by slice illumination. **G)** Non-triggered, but random illumination periods significantly reduce SWR initiation for the duration of the pulse (red). Modified from Schlingloff et al. [2014] with permission from the Society for Neuroscience.

of this correlation could be mediated by a synaptic depression mechanism at the level of the synapses from PV^+ basket cells to pyramidal cells. Indeed, they tested how the ratio between the strengths of two trains of evoked inhibitory postsynaptic currents (IPSCs) correlates with the interval between the two stimulations. The first evoked train induced the depression at the PV^+BCs-to-pyramidal cells synapses. Longer intervals (i.e., longer recovery time from depression) resulted in larger recovery of the IPSC (defined as the ratio of the first IPSC in the second train with respect to the first IPSC in the first train). The recovery time course was very similar to the relation between IEI and following SWR amplitude (see Fig. 1.9B), suggesting that the two phenomena could be related, and that PV^+ basket cells-induced depression could play an important role in regulating the timing of successive SWRs. Additionally, the authors optogenetically activated PV^+ cells to elicit SWRs. They found a similar correlation between the amplitudes of the evoked SWRs and the intervals between the time of stimulations and the last spontaneously occurred SWR (see Fig. 1.9C).

Surprisingly, Chenkov [2017] found no significant correlation when analyzing the amplitude of a SWR event and the length of the following IEI. In this dataset (20 CA3 slices), the author confirmed the existence of a significant correlation between the amplitude of a SWR event and the length of the preceding IEI, but showed that only a weak correlation (correlation coefficient $r < 0.2$) could be observed (in 9/20 slices) between SWR amplitude and following IEI (see Fig. 1.9D). The interpretation of these results remains an open question.

Finally, Schlingloff et al. [2014] suggested that a combination of stochastic and refractory elements could explain the dynamics of SWR occurrence. Indeed, the IEI distributions consistently revealed the presence of a long, approximately exponential tail for large IEI, suggestive of a random component of SWR initiation (see also [Mizuseki and Buzsáki, 2013] and Sec. 1.1.3.2). Similar IEI distributions have been reported *in vivo* [Axmacher et al., 2008; Sullivan et al., 2011]. Additionally, the IEI distributions were characterized by a refractory period lasting ∼100-200 ms, during which no SWR occurred. This result, confirmed by Kohus et al. [2016]; Jiang et al. [2018]; Hunt et al. [2018] for both spontaneous and evoked events, suggests that a refractory mechanism might prevent the occurrence of successive SWRs right after the termination of an event. In light of the results presented by Kohus et al. [2016], it is tempting to conclude that the dynamics of spontaneous and evoked SWR events could be influenced by a PV^+-basket cells-mediated depression mechanism.

1.2.4 How to model SWR generation

The aim of this section is to provide a theoretical framework to the model I am proposing in this thesis. To this end, I review the existing literature on computational models of SWR generation and point out recent experimental findings so far neglected by existing models.

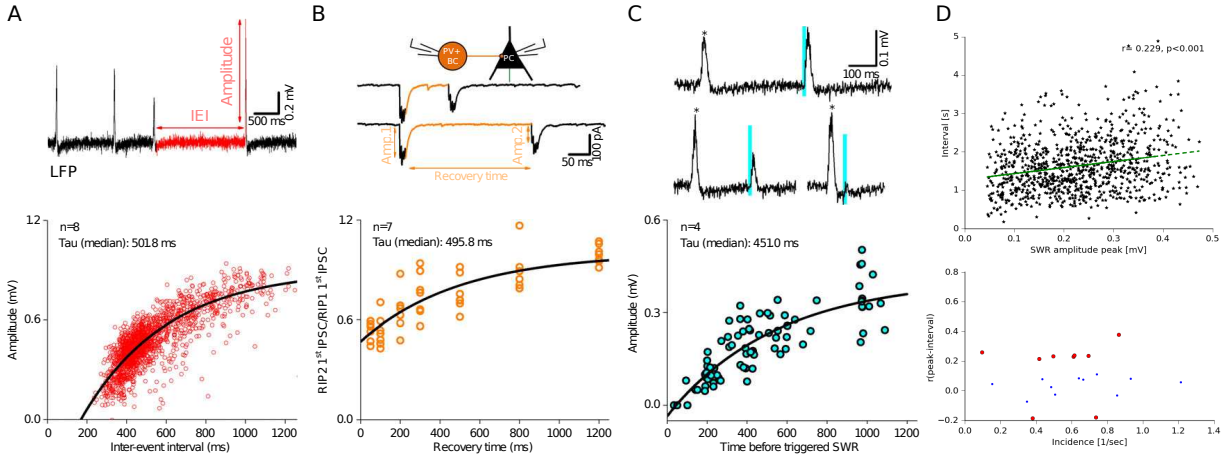

FIGURE 1.9: **Dynamics of spontaneous and evoked SWRs. A)** Spontaneous SWRs recorded *in vitro* in CA3 display a strong correlation between the amplitude of a SWR and the length of the previous IEI. Note the existence of a refractory period (∼200 ms) where no SWRs emerge. **B)** Short-term synaptic depression is induced by repetitive stimulation of a PV^+ basket cell, and IPSCs are measured at a connected pyramidal cell. The ratio of successive IPSCs depends on the interval between successive stimulations (recovery time). See text for details. **C)** Optogenetic stimulation of PV^+ cells successfully evokes SWRs. The dynamics between the amplitude of evoked SWR events, and the difference between time of stimulation and last spontaneous SWR has a similar time course to A) and B). A) - C) are adapted from Kohus et al. [2016] with permission from John Wiley and Sons. **D)** Upper plot: little correlation is present between SWR amplitude and following IEI (Pearson's correlation coefficient is $r = 0.229$). Events from a recording in a CA3 slice are shown. The lower plot presents an overview of all analyzed slices and shows that the correlation does not depend on the SWR incidence in the slice. Red dots depict significant correlations ($p < 10^{-3}$), blue dots non-significant correlations. Adapted from Chenkov [2017].

In this work, I primarily focus on the modeling of the sharp wave component of CA3 SWR events. Ripples are stronger in CA1, and have been shown to emerge independently during development [Leinekugel et al., 2002] and to be uncoupled to CA3 ripples [Sullivan et al., 2011]. Hence, many models have focused on modeling CA1 ripples as independently emerging phenomena. All these ripple models assume an excitatory drive coming from the CA3 area (a sharp wave), but do not explicitly model the SW itself. An additional goal of this work is to formulate a model that can explain the recent experimental findings on the role of interneurons in regulating SWR occurrence presented in Sec. 1.2.3. Clearly, modeling the sharp wave component is of crucial importance in this context. As briefly discussed in Sec. 1.2.2, previous hypotheses on SWR generation have assumed a strong, likely overestimated, recurrent connectivity in the network (see Sec. 1.2.4.1); therefore, it is important to propose alternative models that do not rely so heavily on this assumption. Finally, in the context of memory consolidation, we regard the sharp wave as a mechanism that, by increasing the excitability of (pyramidal) cells, provides a window of opportunity for the reactivation of stored memory traces. In this context, ripple oscillations are likely instrumental to compress the replay to a timescale favorable for plasticity, but I believe that the contribution of the sharp wave should not be neglected.

1.2.4.1 Overview of existing models

Early computational models that aimed at explaining the generation of the sharp wave component focused on the recurrency in the CA3 region as the main feature enabling this region to locally generate population events (SW). In Shen and McNaughton [1996], for example, population bursts emerge randomly in a network of excitatory neurons with place cell-like properties. Neurons coding for overlapping place fields are connected, resulting in an average connectivity of 30%. In Omura et al. [2015], a log-normal distribution of weights among excitatory cells (average recurrent connectivity is 10%) can reproduce the lognormal properties of SWRs reported in Mizuseki and Buzsáki [2013] (see Sec. 1.1.3.2 for details). The network includes inhibitory neurons to avoid activity explosion and a multi-timescale adaptive threshold neuron model [Jolivet et al., 2008; Kobayashi et al., 2009] to generate spike bursts.

Other models, as Dur-e Ahmad et al. [2012]; Taxidis et al. [2012, 2013], made use of single cell bursting as a key mechanism driving population events in CA3. In Dur-e Ahmad et al. [2012], single compartmental excitatory neurons with spike frequency adaptation exhibit bursting behavior when connected in an all-to-all coupled network. The authors claim that this mechanism could support sharp-wave-like events at the population level. Taxidis et al. [2012, 2013] used conductance-based models of pyramidal cells (with 10% recurrent connectivity) and interneurons to generate SWs in CA3, that propagate to CA1, where also ripple oscillations are generated. SWR events are terminated due to pyramidal cells afterhyperpolarization induced by inhibitory neurons. One unrealistic feature of the model is that SWR occurrence is much higher than experimentally observed (i.e., in the theta range, ∼10 Hz, see Sec. 1.2.1 for a comparison with data). Additionally, the model is unable to explain how ripples are generated in the CA3 area (possibly because of the lack of recurrent connectivity among CA3 interneurons). These models were inspired by earlier work by Traub and Wong [1982], who proposed the bursting mechanism as a possible source of epileptic activity, which is considered a pathological deviation from the SWR state.

In addition to models addressing the question of sharp wave generation, a large body of work studied the emergence of oscillations in the ripple range (∼150-250 Hz). Even though ripples are not the main focus of this thesis, I review here existing theories on the emergence of these high-frequency oscillations, to complete the overview on SWR models. Broadly speaking, models of ripple oscillations can be divided in three groups. Models in the excitation-first group propose that excitatory neurons are the source of the high-frequency oscillations, possibly via axonal gap junctions between pyramidal cells [Draguhn et al., 1998; Schmitz et al., 2001; Traub and Bibbig, 2000; Traub and Schmitz, 2012], or due to the interplay between chemical synapses and a supralinear dendritic integration mechanism [Memmesheimer, 2010; Jahnke et al., 2015]. The second class of models includes inhibition-first models, which propose that ripple oscillations are generated by the recurrency in the inhibitory network [Buzsáki et al., 1992; Ylinen et al., 1995; Whittington et al., 2000; Brunel and Wang, 2003; Taxidis et al., 2012; Schlingloff

et al., 2014; Malerba et al., 2016; Donoso et al., 2018]. Somewhat intermediate between the two classes, the third group of models postulates the importance of an interaction between excitatory and inhibitory cells for the emergence of ripple oscillations [Brunel and Wang, 2003; Geisler et al., 2005; Stark et al., 2014].

In recent years, more complex models have been proposed, that attempt to simultaneously model the emergence of SWRs (i.e., both sharp wave and ripple components) in CA3 and CA1, and the sequence learning and subsequent reactivation during SWR events [Cutsuridis et al., 2010; Vladimirov et al., 2013; Romani and Tsodyks, 2015; Jahnke et al., 2015; Malerba and Bazhenov, 2019]. Cutsuridis et al. [2010] modeled sequence encoding during theta oscillations and replay in a CA1 network with a multi-compartmental model for place cells and multiple interneuron types. The model focuses on the reactivation of sequences and SWRs are not explicitly modeled. Vladimirov et al. [2013] showed that a network in which ripples are generated by axonal gap junctions among pyramidal cells is capable of replaying activity in both a forward and backward manner. Romani and Tsodyks [2015] used short-term synaptic depression and excitatory recurrent connectivity to model the creation and reactivation of sequences during theta and SWR states. Jahnke et al. [2015] and Malerba and Bazhenov [2019] presented comprehensive models of CA3 and CA1 oscillations and sequence replay, where SWRs emerge due to nonlinear dendritic interactions, and recurrent excitatory connectivity, respectively.

Other models addressed the question of sequence replay, not necessarily linked to SWR generation. For example, Azizi et al. [2013] showed that a continuous attractor network with spike frequency adaptation can be used to model preplay of multiple spatial environments. Finally, Chenkov et al. [2017] addressed the question of sequence replay in a network (mimicking the CA3 area) where excitation and inhibition are balanced, and replay occurs by subsequent activation of feed-forwardly connected cell assemblies.

An interesting hypothesis on SWR generation, recently presented in Levenstein et al. [2019], suggests that SWRs emerge as deviation from a bistable system, where both SWR and non-SWR states coexist. The model includes a firing rate adaptation mechanism that regulates the termination and occurrence of successive SWRs. The authors hypothesize that the SWR generation mechanism shares many features with the mechanism responsible for the emergence of UP and DOWN states in cortical regions. The idea of bistability to model UP and DOWN states is not new [Mejias et al., 2010; Jercog et al., 2017; Tartaglia and Brunel, 2017], but it was not previously used to model SWRs.

1.2.4.2 Introducing a network to explain the effect of PV^+ cells activation on SWRs

Despite the large variety of models for SWR generation presented in the previous section, none of these models can account for the experiments presented in Schlingloff et al. [2014]; Kohus et al. [2016]. In fact, it remains an open question how the optogenetic activation of PV^+ interneurons can result in an increase of excitation in the network (i.e., it generates a SWR, during which pyramidal cells increase their firing). Surprisingly, even the model

presented in Schlingloff et al. [2014] does not attempt to model these results, and solely focuses on the mechanisms for ripple generation.

A main limitation of all the models presented in Sec. 1.2.4.1 (with the exception of the model by Cutsuridis et al. [2010]) is that they include a single interneuronal population (if at all). In a 2d model comprising one excitatory and one inhibitory population, the excitation of the inhibitory population results in a decrease (and not an increase) of excitatory cells firing, at odds with experimental results. For this reason, I propose an alternative model that includes pyramidal cells and 2 different types of interneurons. In this scenario, the activity of pyramidal cells can increase upon PV^+ cells stimulation because these cells connect to another class of interneurons, that are inhibited by PV^+ cells and thus release the inhibition onto the pyramidal cells. In other words, a *disinhibition* mechanism accounts for the increase in pyramidal cells firing when PV^+ interneurons are stimulated. The development of this model, which has been conceptually introduced in Chenkov [2017], constitutes the core theme of this thesis.

1.3 Outline

The thesis is structured in 5 chapters. In Chapter 2, I introduce a spiking model tuned to reproduce SWRs in an *in vitro* setting. I constrain the model on experimental results and show that it can reproduce a number of experimental features reviewed in Sec. 1.2.3. Specifically, SWRs can emerge in the network spontaneously, or driven by cells stimulation as reported in Schlingloff et al. [2014]; Bazelot et al. [2016]. The dynamics of spontaneous events reproduces experimental results on the relation between the timing of successive SWRs and their amplitude [Kohus et al., 2016; Chenkov, 2017], and supports the hypothesis that a PV^+ basket cells-mediated synaptic depression mechanism could mediate SWR occurrence. Furthermore, this minimal model can be extended to include additional features observed in experiments (as alternative plasticity mechanisms and pharmacological modulation of SWRs). The model predicts the existence of a class of cells that are anti-modulated with respect to SWRs, and suggests that their inactivation is sufficient to initiate a SWR event.

In Chapter 3, I present a mean-field approximation of the spiking model amenable to analytical treatment, that allows us to better understand the network dynamics. I take advantage of the existence of a perturbed bistable configuration in the spiking model to develop a method to analytically predict under which conditions the model displays the desired behavior. This method can be used to refine the spiking model as more experimental details about the microcircuit become available, and can be extended to study other types of bistable disinhibitory networks.

To the best of my knowledge, Chapter 4 presents the first evidence for the existence of a class of CA3 interneurons that are anti-modulated with respect to SWRs. While many details about the identity of the cells remain to be clarified, the recordings convincingly show cells with the predicted firing behavior. To uncover a possibly causal involvement of these anti-SWR cells in the mechanism of SWR generation, I investigate, both in

the spiking model and *in vitro*, the effect of single cell manipulation on the network dynamics. The simulation results suggest that modifying the firing activity of a single anti-SWR cell could have a network-wide effect, but more experiments are needed to confirm whether this holds true in the biological network.

Finally, Chapter 5 concludes the thesis, framing this work in the broader context of cortico-hippocampal memory consolidation, and discussing possible drawbacks of the model and future directions of research.

2 | Simulating SWRs in a spiking network

2.1 Introduction

The aim of this chapter is to define a model capable of explaining how SWRs can be generated in a hippocampal slice. In what follows, I focus on the area CA3 of the hippocampus, which is considered to be the main generation site of SWRs [Maier et al., 2003; Ellender et al., 2010], and I use (mostly *in vitro*) results on rodents as the privileged reference for relevant experimental findings the model should reproduce. The main ingredients on which the model is built have been summarized in Sec. 1.2.3.

To create a simple, but biologically realistic model for SWR generation, we have to restrict ourselves to a set of relevant features we would like the model to reproduce. If the model captures the most important elements of the SWR generation mechanism, it might then generalize across areas and species.

I define the model structure in Sec. 2.2 and show in Sec. 2.3 that the model can be tuned to exhibit both spontaneous and evoked SWRs. In particular, in Sec. 2.3 I focus on the importance of the relative strength of converging network pathways to construct networks that show the desired behavior. I use this knowledge to create a network that accounts for the paradoxical effect of PV^+ cells stimulation in SWR generation [Schlingloff et al., 2014; Kohus et al., 2016]. Furthermore, in Sec. 2.4 I show that the SWR dynamics in the spiking model closely matches experimental results on the termination and occurrence of SWR events. To increase the biological realism, the proposed model can be enriched with additional mechanisms (for example, auxiliary short-term synaptic dynamics and pharmacological modulation), which are discussed in Sec. 2.5. Given the complex structure of the network and the large number of parameters involved, it is difficult to unveil the important mechanisms underlying the network behavior. In Sec. 2.6, I show that the spiking network is in a perturbed bistable configuration. The concept of network bistability can be used to better understand the important features of the network structure and behavior, as presented in Chapter 3. Finally, limitations of the model and further extensions are discussed in Sec. 2.7.

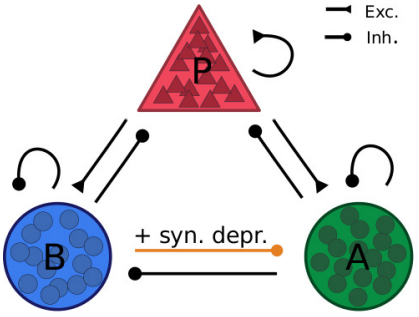

FIGURE 2.1: **Network structure.** The network model comprises a population of pyramidal cells (P) and two groups of interneurons (PV^+BCs and anti-SWR cells, B and A, respectively). Arrows ending with a triangle depict excitatory connections, arrows ending with a circle depict inhibitory connections. The connection from PV^+BCs to anti-SWR cells includes a short-term synaptic depression mechanism.

2.2 Setting up the model

As introduced in Sec. 1.2.4.2, I consider a network with three distinct neuronal populations: one class of excitatory neurons (pyramidal cells, P in what follows) and two classes of inhibitory neurons: PV^+ basket cells (PV^+BCs, called B in the model) and a class of yet unidentified anti-SWR cells (A). I model neurons as conductance-based leaky integrate-and-fire (LIF) units, recurrently connected as depicted in Fig. 2.1.

Conductance-based neurons have been introduced by Hodgkin and Huxley [Hodgkin and Huxley, 1952] and allow for an accurate description of the effect of synaptic interactions on the cell's conductance. At the same time, LIF neurons [Lapicque, 1907; Hill, 1936; Stein, 1967] are one of the simplest models to describe the spiking behavior of a cell, as only the subthreshold behavior of the membrane potential is modeled. Every time the membrane potential reaches a given threshold, a spike is emitted and the membrane is set to a reset voltage value. Due to their simplicity, LIF neuron models have been extensively used to describe information processing in the brain (see e.g. [Gerstner et al., 2014] for a comprehensive review). The choice of combining LIF neurons with conductance-based synapses is discussed in more detail in Sec. 2.7.

The existence of a second class of interneurons in the network (the A cells) allows to explain the experimentally observed increase of pyramidal cell firing upon PV^+ cells activation in terms of a *disinhibition* mechanism: when B cells are activated, A cells are inhibited, and thus release the inhibition onto P cells. This interaction can result in an increased firing of P and B cells and a decreased firing of A cells. We interpret this pattern of activation, in which P, B, and A simultaneously change their firing rates (from low activity to high activity for P and B cells, and from high activity to low activity for A cells), as a signature of a SWR event. Note that this firing behavior is in accordance with experimental results on the activity of hippocampal pyramidal and PV^+ basket cells during SWR events (see Sec. 1.2.3). In the network, the A cells play an important role in controlling the firing of P cells in periods of non-SWR activity, and they become inactive as a result of the increased activity of B cells. The existence of a

class of interneurons tonically active in non-SWR periods and almost silent during SWR events is a prediction of the model, and is discussed in more detail in Chapter 4.

To reproduce the features of the microcircuit present in a rat CA3 slice, I choose cell numbers to match experimentally estimated quantities. This results in 8200 P (pyramidal cells) and 135 B cells (PV^+ basket cells), as explained in detail in Sec. 2.8.2. Additionally, I include 50 anti-SWR cells in the network. Neurons are randomly connected with connection probabilities that have been tuned to reproduce, if available, experimental connectivity results in CA3. At this point, we have no constraints on the connections from and to anti-SWR cells. I tune these connections such that the population firing activity of P and B cells in non-SWR periods and during SWR events is close to experimental values. More details on the choice of parameters and on constructing the network are provided in Secs. 2.8.2 and 2.8.3.

All connections in the network are static (but see Sec. 2.5.1 for an extension), with the exception of the disinhibitory connection $B \to A$. I equip this connection with a short-term synaptic depression mechanism, which is progressively reducing the efficacy of the $B \to A$ synapses as the B cells' firing rates increase. This mechanism is inspired by the results of Kohus et al. [2016], who showed that a PV^+ basket cells-mediated depression mechanism could influence the timing between successive SWR events *in vitro*. In the model, synaptic depression at the $B \to A$ synapses is responsible for the termination of a SWR event, besides influencing the timing and incidence of SWRs.

2.3 SWRs can be generated in a CA3-like spiking network

We can use the features introduced in the previous section to define a spiking network with the desired characteristics. Figure 2.2A shows that the network can be tuned to display spontaneously occurring SWR events (network parameters are listed in Tables 2.1, 2.2, and 2.3). The network has been initialized to start in a non-SWR state, characterized by low activity of P and B cells ($P \approx 2$ spikes/s i.e., spikes per second, and $B \approx 1$ spikes/s), and tonic activity of A cells ($A \approx 12$ spikes/s). In this setup, a SWR event can emerge, characterized by high activity of P and B cells ($P \approx 43$ spikes/s and $B \approx 92$ spikes/s) and almost silent A cells ($A \approx 1$ spikes/s). The initiation of a SWR event is triggered by the intrinsic fluctuations present in the network, caused by the random connectivity and the finite size of the network. As the activity of B cells increases, the values of the synaptic efficacy of the connection $B \to A$ decrease (orange trace in Fig. 2.2 is the mean synaptic efficacy e_{IJ}). As a result, the A cells receive progressively less inhibition from the B cells and thus increase their firing rate and, consequently, their inhibitory drive onto P cells. Thus, the SWR event terminates, and the population firing rates are reset to their non-SWR values. The synaptic efficacy variables have a recovery time constant of $\tau_D = 250$ ms (see Eq. (2.3)), and synaptic efficacy is thus restored (as shown by the orange trace in Fig. 2.2A) more slowly than the population activities.

The last row of Fig. 2.2A shows that SWR events enclose ripple-like oscillations with a peak frequency of 135 Hz (peak of the power spectrum). The existence of ripple oscillations is in line with CA3 *in vitro*-recorded SWR events, even though ripple frequency seems higher in the slice (210 ± 16 Hz, see Sec. 1.2.1).

Additional to spontaneous SWRs, Ellender et al. [2010]; Schlingloff et al. [2014]; Stark et al. [2014]; Bazelot et al. [2016]; Kohus et al. [2016] have shown that SWR events can be triggered by cell stimulation. Can the model reproduce these results? Fig. 2.2B shows that injecting a depolarizing current to a fraction of P cells results in an event that is qualitatively similar to a spontaneous SWR event. Similarly, activating B cells (Fig. 2.2C) by injecting a depolarizing current to a fraction of cells results in a SWR event, as reported in Schlingloff et al. [2014]; Kohus et al. [2016]. Finally, the model predicts that injecting a hyperpolarizing current to the A cells results in the generation of a SWR event comparable to spontaneous events (Fig. 2.2D).

To better understand the involvement of the different populations, Fig. 2.3 displays histograms of unit firing in the SWR events presented in Fig. 2.2. The majority of P cells fire 3 or 4 spikes per event, and only around 5 cells (out of the population of 8200 cells) are silent over the whole duration of a SWR. This result is in disagreement with experiments reporting that pyramidal cells fire, on average, one spike per event [Ellender et al., 2010; Bähner et al., 2011; Hájos et al., 2013], and that only a subset of cells is entrained per event [Ylinen et al., 1995; Ellender et al., 2010; Mizuseki and Buzsáki, 2013]. I discuss possible explanations for this discrepancy in Sec. 2.7. The central plot of Fig. 2.3 shows that most of the B cells fire 5-9 spikes per event, in line with experimental evidence [Hájos et al., 2013; Schlingloff et al., 2014]. Finally, a large portion of the A cells is silent during SWRs, as predicted by the fact that A cells' inhibition is needed to release the inhibition of P cells. Overall, the figure shows that the involvement of cells in SWR events is largely independent on the initiating mechanism (spontaneous vs. current-induced).

2.3.1 The relative strength of converging pathways influences the emergence of SWRs

The results of current stimulation presented in Fig. 2.2 crucially depend on the connectivity in the network: if, for example, the connection $B \to P$ was very strong, activating B cells could result in a decrease of P cells' firing, and not in a disinhibition-mediated increase of firing. Similarly, A activity could increase upon activation of P cells if the connection $P \to A$ was too strong. These considerations can be summarized in a series of requirements on the strength of converging pathways in the network:

1. Pathway $P \to B \to A$ should be stronger than $P \to A$. This is required to guarantee that the activation of P decreases A firing (i.e., it activates the disinhibition mechanism).
2. Pathway $P \to B$ should be stronger than $P \to A \to B$. This is required to guarantee that the activation of P increases B firing.

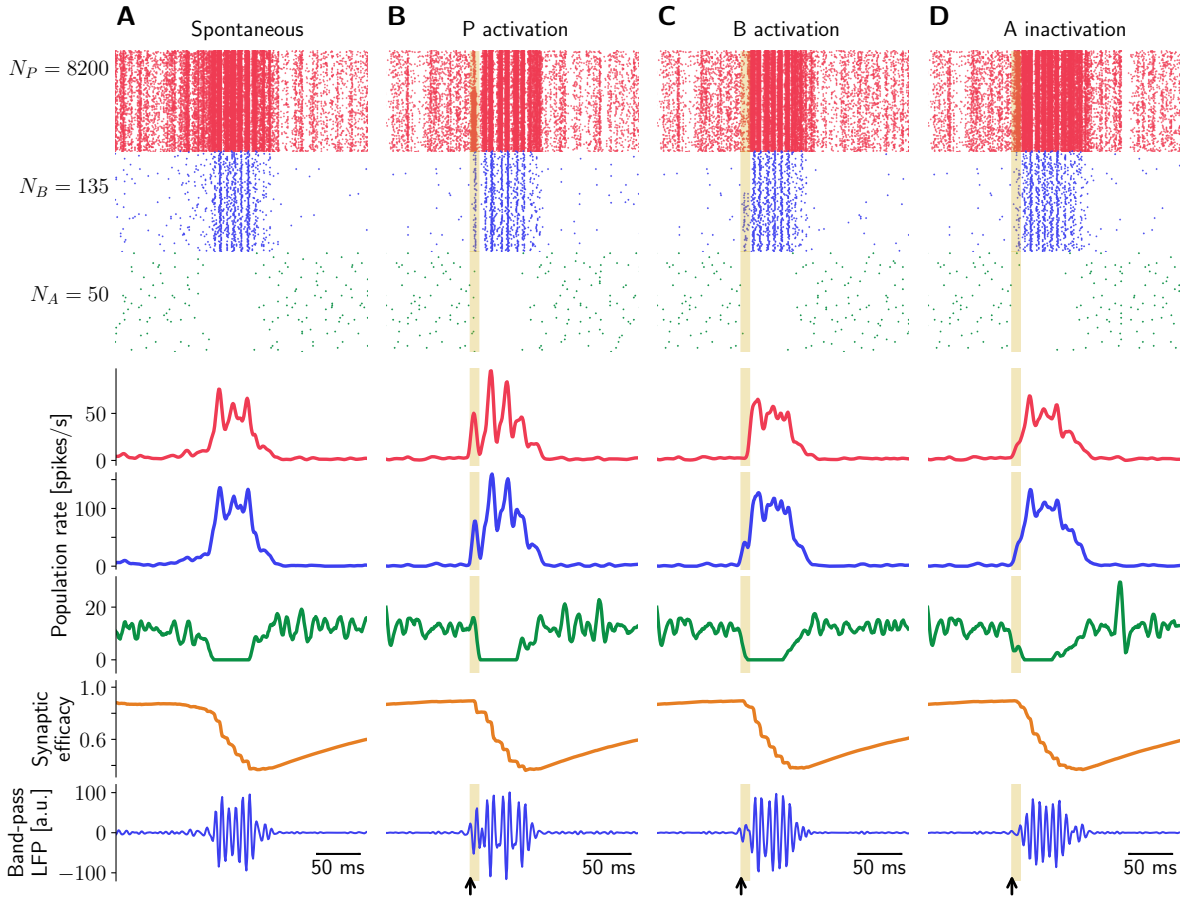

FIGURE 2.2: **SWR events can be generated in the spiking network.** Each column describes one simulation. **A)** Spontaneous events can be generated in the network. Displayed are raster plots for P, B, and A cells (red, blue, and green, respectively, each line is one neuron), averaged population firing rates, the time course of the synaptic depression mechanism, and the band-pass filtered component of the LFP signal (see Sec. 2.8.4). **B), C)** A fraction of P and B cells are stimulated with depolarizing current for $T = 10$ ms (yellow area). The stimulation elicits events comparable with the spontaneous ones, in agreement with experimental results. **D)** A fraction of A cells is stimulated with hyperpolarizing current. The stimulation elicits events comparable with the spontaneous ones; this is a prediction of the model.

3. Pathway $B \to A \to P$ should be stronger than $B \to P$. This is required to guarantee that the activation of B increases P firing.
4. The two pathways $B \to A$ and $B \to P \to A$ collaborate to decrease the activation of A upon activation of B.
5. Pathway $A \to P$ should be stronger than $A \to B \to P$. This is required to guarantee that the inactivation of A increases P firing.
6. The two pathways $A \to B$ and $A \to P \to B$ collaborate to increase the activation of B upon inactivation of A.

In formulating these requirements, we are implicitly incorporating the recurrence of the populations (e.g., the recurrent A connection in the pathway $B \to A \to P$). Furthermore, we are neglecting any temporal structure in the network. These requirements need to be taken into account when constructing the network, to guarantee that SWR

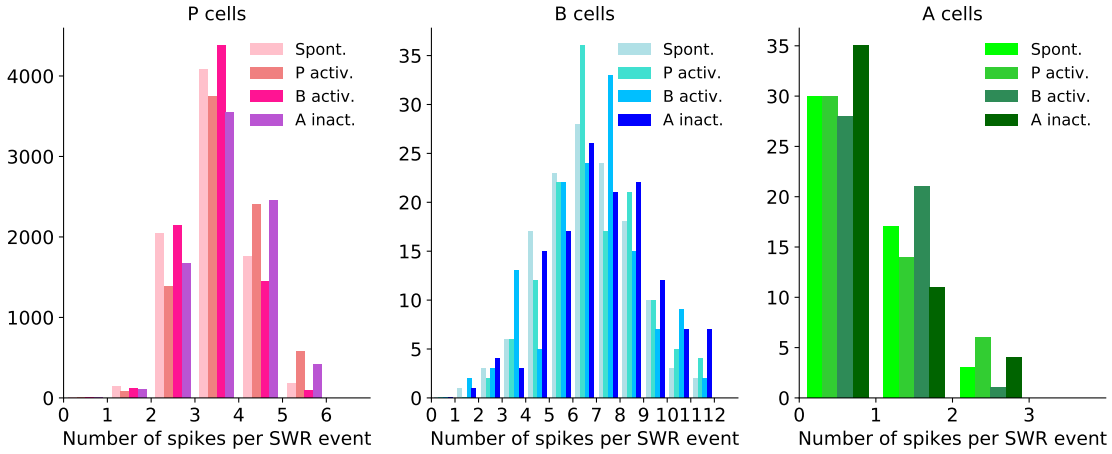

FIGURE 2.3: **Cell participation to SWR events.** The histograms show the number of spikes per SWR event of each population (P, B, and A cells, from left to right). A SWR event is defined here as a 100 ms-long interval of larger-than-baseline activity of P cells. The different colors correspond to the SWR events presented in Fig. 2.2, i.e., in the conditions of spontaneous events, and SWRs induced by activation of P and B cells, or inactivation of A cells (from light to dark colors, see legend). Note that only very few P cells ($\sim$5 cells) remain silent for the whole duration of a SWR (not visible in plot).

events evoked in the network are similar to experimentally recorded SWRs. Specifically, in Sec. 2.8.3 I explain how these conditions can be used to tune the spiking network to show the desired behavior. Additionally, I show in Chapter 3 how requirements 1, 2, 3, and 5 can be formalized and enforced to construct networks with the aforementioned characteristics.

2.4 Features of spontaneous and evoked SWRs match experimental results

Next, I test whether the spiking network can reproduce the results presented in Sec. 1.2.3 relative to the dynamics of SWR generation and termination (including the correlation structure across consecutive events). With this intention, I record long simulations (approximately 10 minutes of simulated time) during which SWR events occur spontaneously with an incidence of $\sim$1.3/s.

To be able to better compare the results to experimental SWRs, which are usually recorded in the LFP of *stratum pyramidale*, we need to define a measure for the local field potential in our setting. The exact origin and cellular contribution to the LFP remain elusive, and the subject is a matter of intense debate (see, e.g., [Einevoll et al., 2013]). I decided to approximate the LFP with a low-pass filtered version of the average input current from B cells to P cells. More details about the procedure and its motivation are discussed in Sec. 2.8.4.

A small portion of the raw and low-pass filtered traces are shown in the upper row of Fig. 2.4. Properties of spontaneous SWRs in the network are shown in Fig. 2.4B1. The

IEI distribution is close to experimental results [Schlingloff et al., 2014; Jiang et al., 2018], where the mean IEIs are in the range of 0.5-1 second. The amplitude of the (sharp wave, or SW) event and the full width at half maximum (FWHM) cannot be easily compared to experiments because our measure of the LFP is not the same as the recorded extracellular signal (see Sec. 2.8.4). However, both SW amplitude and duration have been shown to vary over the course of a recording [Ellender et al., 2010; Sullivan et al., 2011; Hofer et al., 2015], similar to the results shown in Fig. 2.4B1. In agreement with Kohus et al. [2016]; Chenkov [2017]; Jiang et al. [2018], we observe a strong correlation (Pearson correlation coefficient $c = 0.57, p < 0.001$) between the event amplitude and the length of the previous IEI, as shown in Fig. 2.4C1 (compare with Fig. 1.9A; as no correlation coefficients are calculated, we rely on visual inspection). Interestingly, the correlation between the event amplitude and the length of the next IEI is low (Pearson correlation coefficient $c = -0.06, p = 0.079$). The lack of correlation between event amplitude and next IEI has been reported by Chenkov [2017] in CA3 *in vitro* data (see Fig. 1.9D).

I additionally consider the case of evoked events, to further compare the simulations with experimental results [Kohus et al., 2016]. In particular, it has been shown that optogenetic drive can elicit SWRs with shorter IEIs than the spontaneous events, but with the same correlation structure between IEI and amplitude discussed in the previous paragraph. Figure 2.4 shows that the spiking network reproduces this behavior. In these simulations, SWRs occur spontaneously, but are additionally triggered by stimulation of B cells (similar to Kohus et al. [2016], see Fig. 1.9C). A short snapshot of the simulation is shown in Fig. 2.4A2. The properties of evoked events are then analyzed, as well as the correlation between the amplitude of an evoked event and the timing to the previous (respectively next) spontaneous SWR. Figure 2.4B2 shows that evoked SWRs are all-or-none events, with IEI distribution, amplitude, and FWHM similar to spontaneous events. Figure 2.4C2 shows the presence of a strong correlation between the amplitude of evoked SWRs and the length of the previous IEI (Pearson correlation coefficient $c = 0.77, p < 0.001$), but not with the next IEI (Pearson correlation coefficient $c = 0.01, p = 0.889$).

These simulations also reveal the existence of a refractory period of $\sim$170 ms following a SWR event (see dashed line in Fig. 2.4C1) during which no new events are generated, in line with Kohus et al. [2016]; Schlingloff et al. [2014]; Jiang et al. [2018] (see Sec. 1.2.3 for details). The length of the refractory period is expected to correlate with the strength of the stimulation, which also explains why evoked events can be triggered at $\sim$65 ms (see dashed line in Fig. 2.4C2), when no spontaneous SWRs emerge. Note, in fact, that the internal fluctuations initiating a spontaneous event can be considered as small input currents from the surrounding network.

In the context of this model, the correlation structure can be explained by the dynamics of the synaptic depression mechanism. The time that has passed since the last event determines how much the depression has recovered, and thus how effective the connection $B \to A$ is. Longer recovery times mean that B cells are effective in inhibiting A cells, which makes the initiation of a new SWR event possible. The recovery of depression explains the existence of a significant correlation between the length of the previous

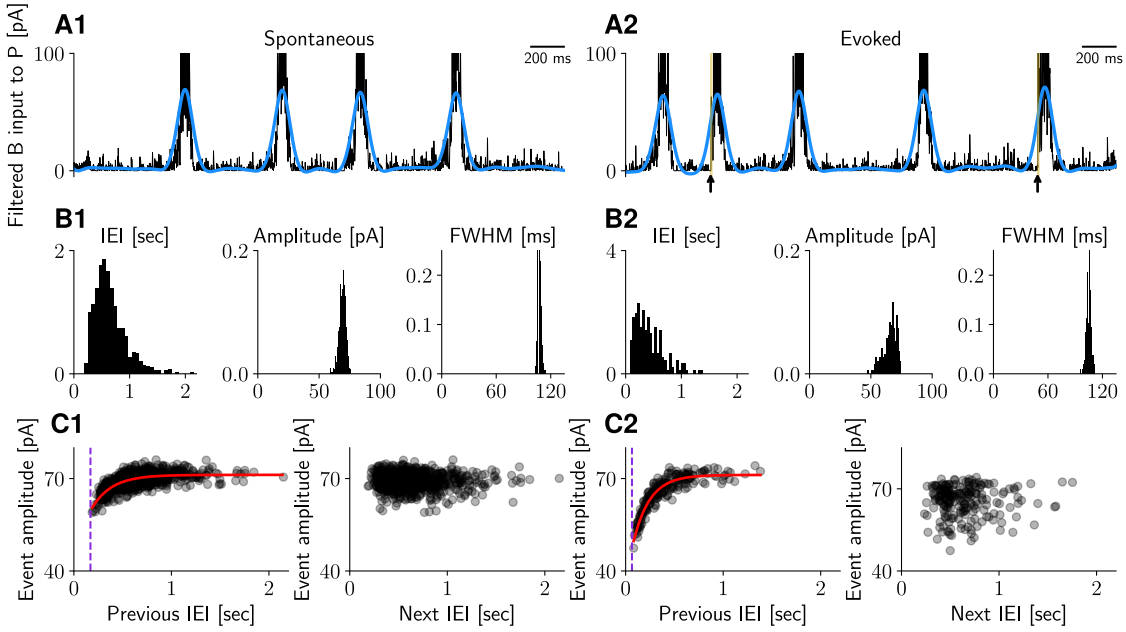

FIGURE 2.4: **Properties of spontaneous and evoked SWRs.** Left: analysis of spontaneous events. **A1)** Mean B input current to P cells (sign-reversed, black trace) and its low-pass filtered version (blue trace). **B1)** Properties of spontaneous SWRs: inter-event-intervals (IEI, distance from end to start of events, where start and end points are calculated at half maximum of the filtered signal), amplitude of filtered events, full width at half maximum (see Sec. 2.8.4 for details). **C1)** Left, strong correlation between event amplitude and previous IEI. Each dot is a pair, and the red line depicts the best fit exponential function. The dashed line indicates a lower bound for the IEI (170 ms). Right, negligible correlation between amplitude of event and length of the next IEI (see text for details). **A2), B2), C2)** Same as A1, B1, and C1, but for events evoked by current stimulation to B cells (current is injected for $T = 10$ ms, black arrows and yellow areas in A2). In C2, the dashed line indicates a lower bound for the IEI (65 ms). Correlation results are in line with experimental observations by Kohus et al. [2016]; Chenkov [2017].

IEI and the amplitude of an event, as well as the existence of a refractory period during which no SWRs are generated. Conversely, whenever the synaptic efficacy decreases below a critical value, the inhibition at the synapses $B \to A$ becomes ineffective, and this induces the termination of an event by restoring the high activity of A cells. The time to the next event is then determined by the recovery of the synaptic efficacy variables. The recovery of depression, but not the amplitude of the previous event, is thus expected to influence the timing to the next spontaneous SWR, suggesting that the correlation between the event amplitude and the length of the next IEI should be low.

2.5 Model extensions

While creating a spiking network, I tried to include a minimal set of components, which are sufficient to reproduce the experimental findings of interest. I am thus undoubtedly neglecting many other phenomena, which might also contribute to the modulation of SWR dynamics. In this section, I focus on two different sets of model extensions: in

Sec. 2.5.1, I study the effect of additional short-term plasticity mechanisms in the network. In Sec. 2.5.2, I model the effect of pharmacological modulation of the synaptic strengths on the SWR incidence.

2.5.1 Additional short-term plasticity mechanisms

In this section, I describe how plasticity mechanisms additional to the $B \to A$ synaptic depression could be included in the model. It is indeed known that, both in hippocampus and neocortex, the synapses from PV^+BCs to pyramidal cells are depressing [Galarreta and Hestrin, 1998; Kraushaar and Jonas, 2000; Szabó et al., 2010; Kohus et al., 2016], as well as those among PV^+BCs [Bartos et al., 2002; Kohus et al., 2016]. Some studies [Galarreta and Hestrin, 1998; Reyes et al., 1998; Losonczy et al., 2002] also report the existence of a pyramidal cells-to-PV^+BCs depression *in vitro*, even though [Pala and Petersen, 2015] could not confirm it *in vivo*.

B $\to$ P synaptic depression. How would additional depression mechanisms modify the network behavior? To answer this question, I test the effect of a $B \to P$ synaptic depression mechanism in the model, and compare the results with the default case (i.e., with the case in which the $B \to A$ synapses are the only plastic connections). For simplicity, I assume that the properties of depression (time decay and plasticity rate) are identical to those of the $B \to A$ depression. This assumption could be valid given that both mechanisms share the same presynaptic population (see Eq. (2.3) in Sec. 2.8.2 for details about how to model synaptic depression).

As a result of the $B \to P$ depression, the B inhibition onto pyramidal cells is reduced. How does this impact the SWR properties in our setup? As in the default scenario with plastic $B \to A$ connection, the depression gets activated during a SWR event, when B cells increase their firing rate. Hence, P cells receive less inhibition while being already active. This suggests, for example, that the population rate of P cells increases while the $B \to P$ depression is on. Figure 2.5 shows simulations of the spontaneous network when both depression mechanisms are active. Panel A shows that the approximated sharp wave signal is lower when the $B \to P$ depression is present. Given that the LFP is defined as a filtered version of the mean B input to P cells (see Sec. 2.8.4 for details), the effect is not surprising. As a result, the population firing rates of both P and B cells increase with added $B \to P$ depression (panel B), as expected. The properties of the approximated sharp wave signal are quantified in Fig. 2.5C. As discussed, $B \to P$ depression decreases event amplitudes, whereas the IEIs remain largely unaffected. The FWHM is slightly lower in the scenario with $B \to P$ depression. However, it is important to keep in mind that the FWHM is intrinsically linked to the event amplitude, thus it can be misleading to compare it across conditions where events have different amplitudes. Finally, panel D shows that the correlation structure of SW amplitude and previous or next IEI stays remarkably unchanged when the $B \to P$ depression is added (Pearson correlation coefficient for case with $B \to P$ depression: amplitude and previous IEI: $c = 0.81$, $p < 0.001$, amplitude and next IEI: $c = 0.04$, $p = 0.204$). Overall, we can

conclude that the network properties are largely preserved when a $B \to P$ depression mechanism is added to the default network with $B \to A$ depression.

Could the $B \to P$ depression mechanism replace the $B \to A$ depression in the network and generate spontaneous events? In a network with $B \to P$ depression alone (i.e., with $B \to A$ connections not plastic), P cells receive less inhibition when they are active (during a SWR), and thus persist in an active state. Hence, the network cannot escape from the SWR state and events do not terminate (simulation not shown). In this sense, the $B \to P$ depression can be thought of as an additional, but not alternative, mechanism to the $B \to A$ depression.

P → A synaptic facilitation. Another prominent plasticity mechanism is the facilitation at synapses connecting pyramidal cells to different types of interneurons [Reyes et al., 1998; Wierenga and Wadman, 2003; Silberberg and Markram, 2007; Pala and Petersen, 2015; English et al., 2017]. In hippocampus, this mechanism has been mostly investigated for O-LM cells [Ali and Thomson, 1998; Losonczy et al., 2002; Böhm et al., 2015]. Because the identity of anti-SWR cells is currently unknown, this property could be interesting to consider in the network. The $P \to A$ facilitation is expected to increase the excitation seen by the A cells when P cells are active, i.e., during a SWR. Thus, this mechanism could help restoring the firing rate of A cells characteristic of non-SWR states. I test the effect of $P \to A$ facilitation comparing the behavior of the default network with the one of a network to which this mechanism is added (for details about the implementation, see Sec. 2.8.2). Figure 2.6 shows an overview of the facilitation effects in the network: the amplitude of the LFP signal is slightly reduced, likely due to a stronger $A \to B$ inhibition caused by more active A cells. Additionally, the IEI distribution is shifted to larger values in the case of added facilitation, possibly because the recovery of both $B \to A$ depression and $P \to A$ facilitation is needed to start an event. Interestingly, the overall correlation structure shows an unchanged trend (Pearson correlation coefficient for case with facilitation: amplitude and previous IEI: $c = 0.48$, $p < 0.001$, amplitude and next IEI: $c = -0.02$, $p = 0.567$). Overall, we can conclude that the network is relatively robust to the addition of a $P \to A$ facilitation mechanism.

Furthermore, it can be shown that the facilitation mechanism alone is sufficient to terminate a SWR event, i.e., it can successfully replace the $B \to A$ synaptic depression. To show this, I simulate a network where the $P \to A$ facilitation is the only plastic mechanism in the network (i.e., the synaptic efficacy of the $B \to A$ connection is clamped at $e_{IJ} = 1$ for the whole duration of the simulation). Figure 2.7 shows that spontaneous events emerge in a network with $P \to A$ facilitation alone (see Sec. 2.8.2 for details). These events can have a much longer duration (as indicated by the FWHM) than the ones in the default network. This phenomenon can be explained by recognizing that, in the network with facilitation, the initiation and termination mechanisms are distinct. Indeed, an event is initiated when fluctuations at B cells are large enough to inhibit the activity of A cells. For this, the $B \to A$ connection needs to be large (synaptic efficacy $e_{IJ} \approx 1$, i.e. almost no depression active). When an event starts, the facilitation $P \to A$

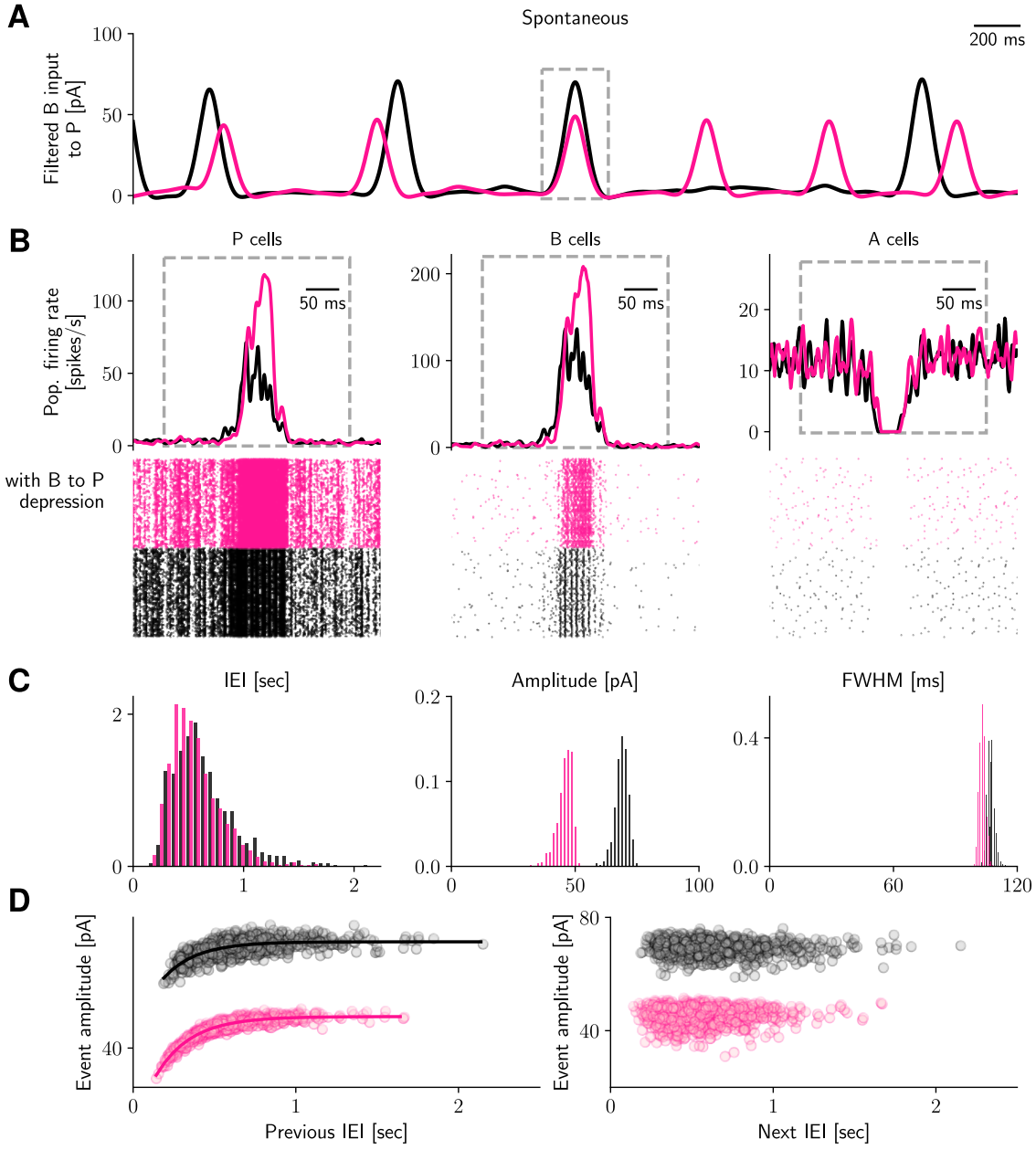

FIGURE 2.5: **Effect of PV$^+$BCs-to-pyramidal cells synaptic depression. A)** Snapshot of spontaneous, low-pass filtered LFP activity in default setting (black, as Fig. 2.4) and in the scenario where $B \to P$ synaptic depression is added (pink). **B)** One event is isolated, and the corresponding population firing rates and cells' raster plots are shown (for P, B, and A cells, respectively). Events are aligned with respect to the peak of the LFP signal. **C)** Properties of spontaneous events in control (black) and $B \to P$ depressed (pink) scenarios. Features are defined in Sec. 2.8.4. **D)** Correlation structure of sharp wave amplitude and previous (left) and next (right) IEI are remarkably similar in the two scenarios. The shift along the vertical axis is caused by the decreased event amplitude in the case with $B \to P$ depression. Solid lines describe best fit exponential functions.

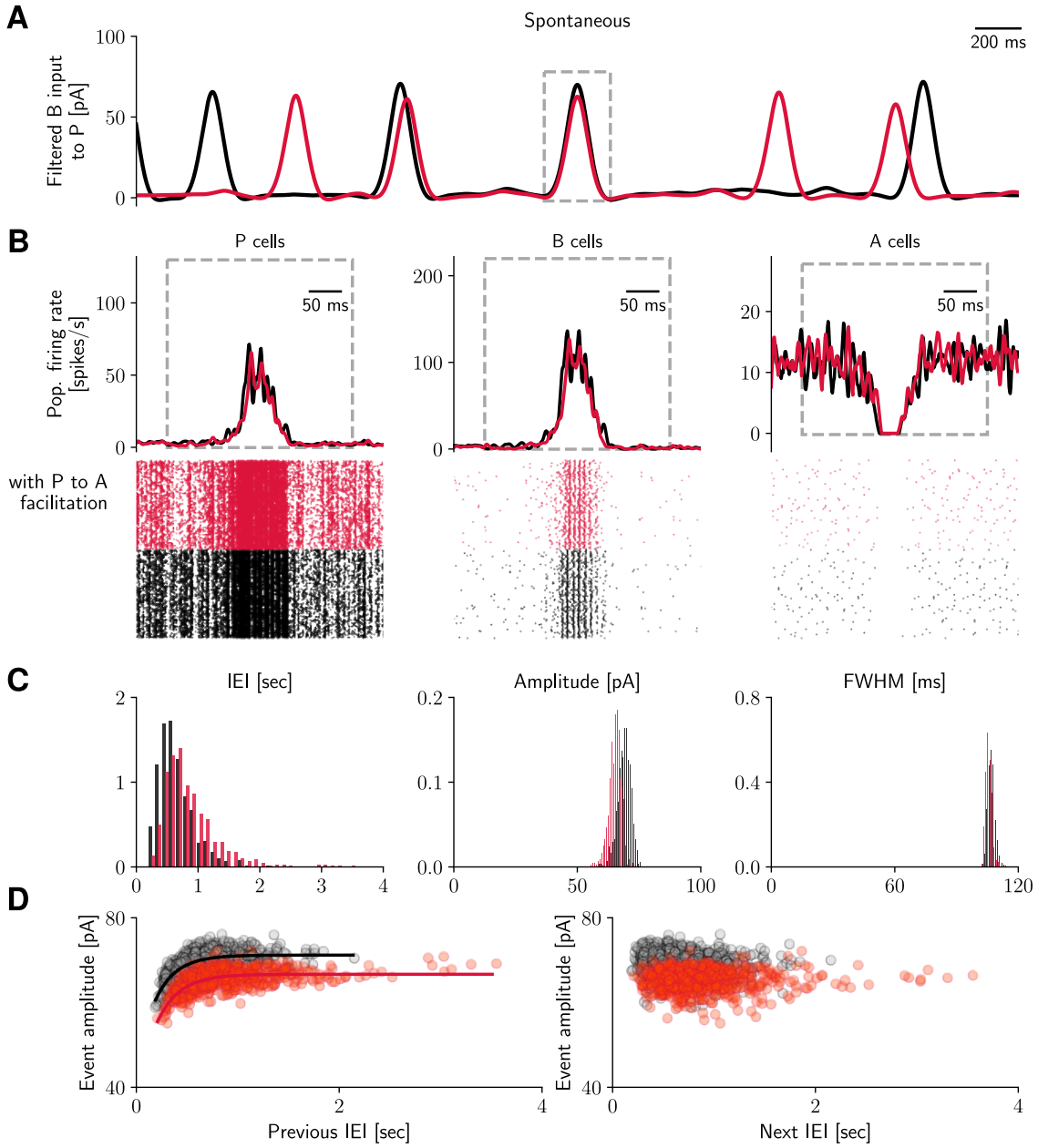

FIGURE 2.6: **Effect of pyramidal-to-anti-SWR cells synaptic facilitation. A)** Snapshot of spontaneous, low-pass filtered LFP activity in default setting (black) and in the scenario where $P \rightarrow A$ synaptic facilitation is added (red). **B)** One event is isolated, and the corresponding population firing rates and cells' raster plots are shown (for P, B, and A cells, respectively). Events are aligned with respect to the peak of the LFP signal. **C)** Properties of spontaneous events in control (black) and facilitation (red) scenarios. Features are defined in Sec. 2.8.4. **D)** Correlation structure of SW amplitude and previous (left) and next (right) IEI are remarkably similar in the two scenarios. The shift along the vertical axis is caused by the decreased event amplitude in the case with $P \rightarrow A$ facilitation. Solid lines describe best fit exponential functions.

gets large and this increases the firing of A cells. During this process, fluctuations in B can still prompt the inhibition of A cells. Thus, the A cells might get partially inhibited again, and a longer time is required for the facilitation to make the A cells fully active again, and thus terminate a SWR event. Conversely, in the default case with only $B \rightarrow A$ depression, the initiation and termination mechanisms are both dependent on the $B \rightarrow A$ connection, preventing the occurrence of long events. In other words, during a SWR event the effect of possible fluctuations in the B cells activity is suppressed by the (depression-driven) lower efficacy of the $B \rightarrow A$ connection, and no new events can be triggered before the depression recovers (not shown).

A remarkable feature of the simulations with only the $P \rightarrow A$ facilitation mechanism is that the network can still reproduce the existence of a strong correlation between event amplitude and previous IEI experimentally observed by Kohus et al. [2016] (Pearson correlation coefficient, amplitude and previous IEI: $c = 0.41$, $p < 0.001$, amplitude and next IEI: $c = 0.08$, $p = 0.020$). This result suggests that the SWR termination mechanism is a main component influencing the existence of the correlation between IEIs and amplitude.

The results presented in this section suggest that multiple plasticity mechanisms could be simultaneously in place in the biological system, and that their interplay might regulate when SWRs are initiated and terminated.

2.5.2 Pharmacological modulation of SWRs by gabazine

The *in vitro* setup allows us to study SWR properties in a targeted manner. A prolific line of research deals with the pharmacological control of SWRs: can drugs be applied to the slice that modify the occurrence and features of SWR events? This approach has been instrumental to better understand the contribution of inhibition to the generation of SWRs. In particular, I focus here on the effect of modifying the dynamics of $GABA_A$ receptors, i.e., channels that conduct chloride ions when activated by GABA (γ-Aminobutyric acid), which is the main inhibitory neurotransmitter of the brain. Typically, the cell membrane gets hyperpolarized due to the induction of an inhibitory postsynaptic potential (IPSP, but possible alternative mechanisms exist, which result in a long-lasting 'tonic' hyperpolarization). Several experiments have tried to alter the inhibitory transmission in the hippocampus to study the effect on SWRs. Papatheodoropoulos et al. [2007] showed that applying low doses of thiopental, a drug that increases the decay time constant of IPSPs, decreases SWR incidence and increases their duration. Furthermore, low doses of diazepam (known to increase the inhibitory peak conductance) increase SWR incidence, whereas high concentrations have an opposite effect [Koniaris et al., 2011].

In addition to studies focusing on enhancing the strength of the IPSPs, several other papers looked at the effects of reducing the efficacy of inhibitory transmission. For this purpose, a widely used drug is gabazine (SR-95531), which is a $GABA_A$ receptor antagonist that binds to the GABA binding site on the receptors and decreases their inhibitory conductance. Given its role in decreasing the strength of inhibition in the

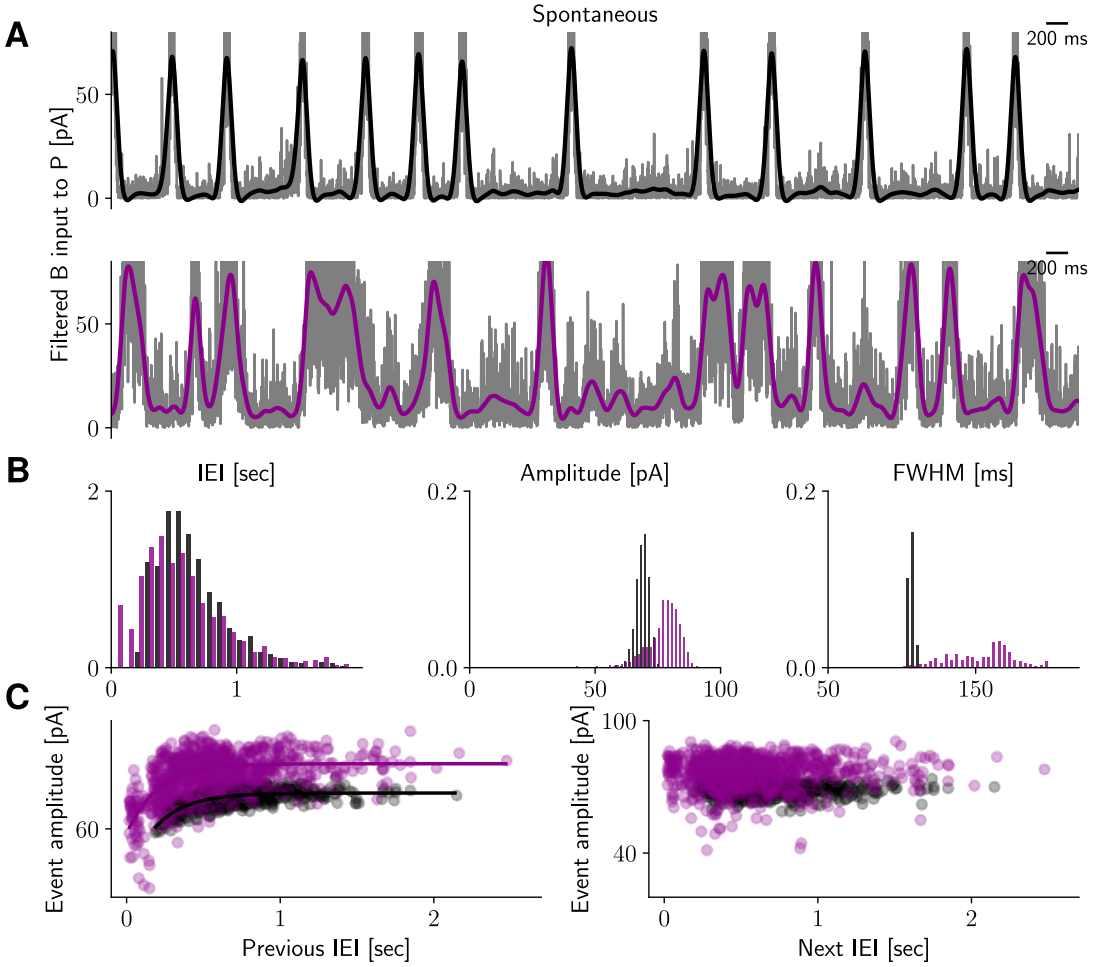

FIGURE 2.7: **Pyramidal-to-anti-SWR cells synaptic facilitation can replace the PV$^+$BCs-to-anti-SWR cells depression in the network. A)** Snapshot of spontaneous, low-pass filtered LFP activity in default setting (black) and in the scenario with $P \rightarrow A$ synaptic facilitation alone (i.e., no $B \rightarrow A$ depression, purple). **B)** Properties of spontaneous events in control (black) and facilitation (purple) scenarios. Features are defined in Sec. 2.8.4. **C)** Correlation structure of SW amplitude and previous (left) and next (right) IEI is preserved in the scenario with $P \rightarrow A$ facilitation alone. Solid lines describe best fit exponential functions. In the simulation with $P \rightarrow A$ facilitation, the reciprocal connections among interneurons are adjusted to yield enough events (see Sec. 2.8.2).

network, it might be tempting to predict that gabazine favors SWR occurrence. However, it has been shown that the application of low doses of gabazine decreases SWR incidence [Nimmrich et al., 2005; Chenkov, 2017], and that higher doses completely abolish SWRs [Maier et al., 2003; Ellender et al., 2010]. These studies also show that SWRs can be replaced by epileptic-like activity in the high gabazine condition (see also Behrens et al. [2007]). Even though these results support the idea of inhibitory neurons playing a pivotal role in SWR generation, in line with the model's hypothesis, a clear interpretation is still lacking. To shed light on the effect of gabazine in regulating SWR events, I study how it affects the SWR dynamics in the spiking network defined in Sec. 2.3.

Uniform reduction of inhibitory conductances. To mimic gabazine application, I reduce all inhibitory conductances by a given fraction and study the effect on the occurrence and dynamics of SWRs. I monitor the network for 20 minutes of simulated time; at $T = 10$ minutes, I 'turn on' gabazine and assume that it instantaneously reduces the strength of all GABAergic conductances (details on the simulations are provided in Sec. 2.8.4). In Fig. 2.8, I show the effect of a 30% reduction of all conductances (which is in line with the estimation on the effect of 0.1-0.2 μM gabazine application in Ellender et al. [2010]). The amplitude of the LFP signal is strongly reduced (panel B); this matches the results in Schlingloff et al. [2014], but is in disagreement with the analysis of experiments in Chenkov [2017], where a small increase in SW amplitude was detected. At the level of populations, B cells strongly increase their firing rates, whereas the firing of P cells is slightly reduced (not shown). Panel C of Fig. 2.8 shows that the correlation structure is largely unchanged with respect to the control condition (correlation coefficients not significantly different, Fisher z-transformation test). Panel D shows that the SWR incidence is only mildly reduced. A lower reduction of inhibitory conductances (-10%) has a smaller effect on the amplitude and a similar effect of SWR incidence. Thus, it seems that these simulations cannot explain the experiments showing a marked decrease in SWR incidence upon application of low doses of gabazine.

To complicate the picture even further, Fig. 2.9 shows that the effect of gabazine can be highly nonlinear. When reducing the strength of all inhibitory conductances by 50%, the gabazine condition shows a strong increase in SWR incidence, which is at odds with the effects described in Fig. 2.8. In the simulation shown in Fig. 2.9, the correlation is stronger in the gabazine condition (panel C, $c = 0.57$ to $c = 0.72$, in control and gabazine condition, respectively, $p < 0.001$), and both P and B cells increase their firing rates (not shown). Note that the SW amplitude is strongly reduced compared to the control condition (panel B); as the conductance strength is reduced further (e.g., the gabazine dose is increased to reduce conductances by 70-80%), the events become indistinguishable from the background activity.

Reduction of selected inhibitory conductances. As little is known about how gabazine affects different types of GABAergic connections [Sieghart et al., 2002], we can speculate that it might have a differential effect on inhibitory synapses. For example, we can assume that gabazine affects the 'disinhibitory' connection $B \to A$ while leaving the other connections unchanged, and test the effect on the SWR incidence. This simulation is shown in Fig. 2.10, where the conductances of the connection $B \to A$ are reduced by 10%. As a result, the SWR incidence drops by 50% (panel D). A stronger reduction of the $B \to A$ conductances completely abolishes SWRs in the gabazine condition, confirming the importance of the disinhibitory connection in SWR initiation. Interestingly, the SW amplitude is only mildly affected (Fig. 2.10B), and both B and P cells reduce their firing rates. Even though the correlation structure is preserved (see Fig. 2.10C), the correlation coefficients are significantly different (from $c = 0.6$ to $c = 0.38$, in control and gabazine condition, respectively, $p < 0.001$). This result contrasts with the

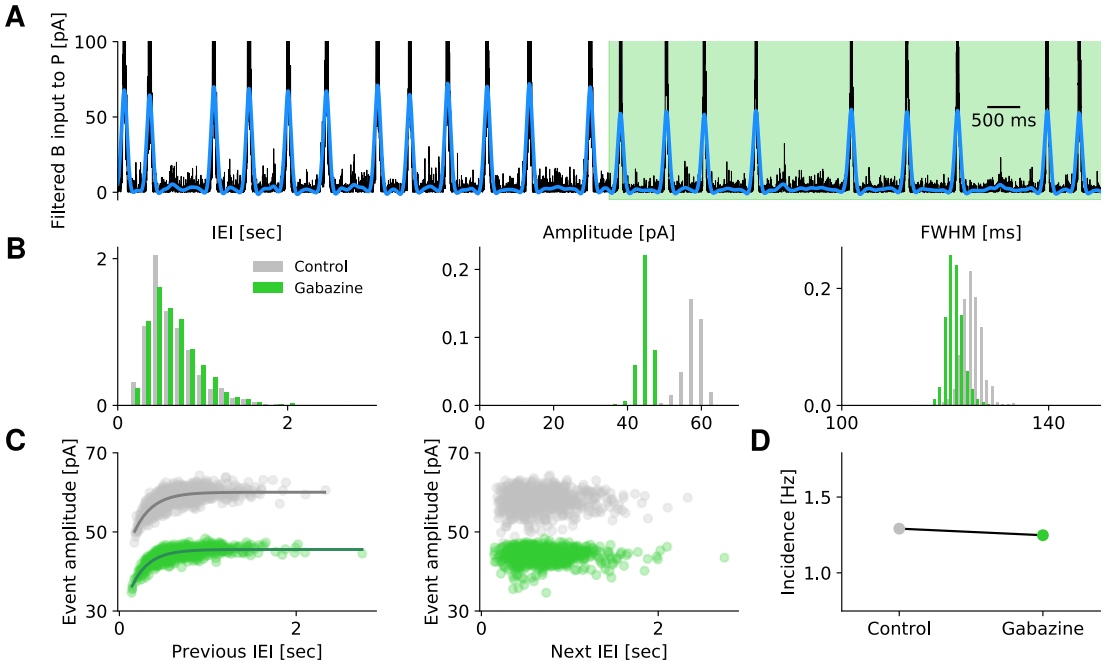

FIGURE 2.8: **Mildly reducing all inhibitory conductances (−30%) has little effect on SWR incidence.** All inhibitory conductances have been reduced by 30% to mimic the effect of low doses of gabazine. **A)** A snapshot of the simulation is shown. Displayed is the LFP signal (average mean B input current to P cells, in black, and the low-pass filtered signal, in blue). The green area depicts the period with reduced inhibitory conductances (gabazine condition). **B)** SWR properties in control condition (gray) and in the gabazine condition (green). Note the amplitude reduction with gabazine. **C)** The correlation structure is largely unaltered by gabazine. Left: the event amplitude plotted against the length of the previous IEI shows a strong correlation in control (as in Fig. 2.4) and gabazine conditions. Right: the correlation is not present when the event amplitude is plotted against the length of the next IEI (as in Fig. 2.4). **D)** Gabazine reduces only mildly the SWR incidence with respect to the control condition.

analysis performed in Chenkov [2017], where the correlation between previous IEI and sharp wave amplitude was found to increase with gabazine in CA3 slices. A similar reduction of SWR incidence can be obtained by selectively reducing the strength of the connections $A \to A$ (and by simultaneously reducing the strength of both $B \to A$ and $A \to A$ connections, not shown). On the other hand, SWR incidence increases when the connections $B \to B$, $B \to A$, or $B \to P$ are selectively weakened (and the other connections are unchanged, not shown). Given the convergence of multiple pathways on each population, it is difficult to understand the combined effect of 'applying' gabazine when different connections are differentially modulated.

In summary, I have shown that the effect of gabazine application can be studied in the spiking network by reducing the strength of the GABAergic conductances. The reduction in SWR incidence observed in Nimmrich et al. [2005]; Chenkov [2017] can be reproduced when the disinhibitory connection ($B \to A$) is selectively targeted by gabazine, but

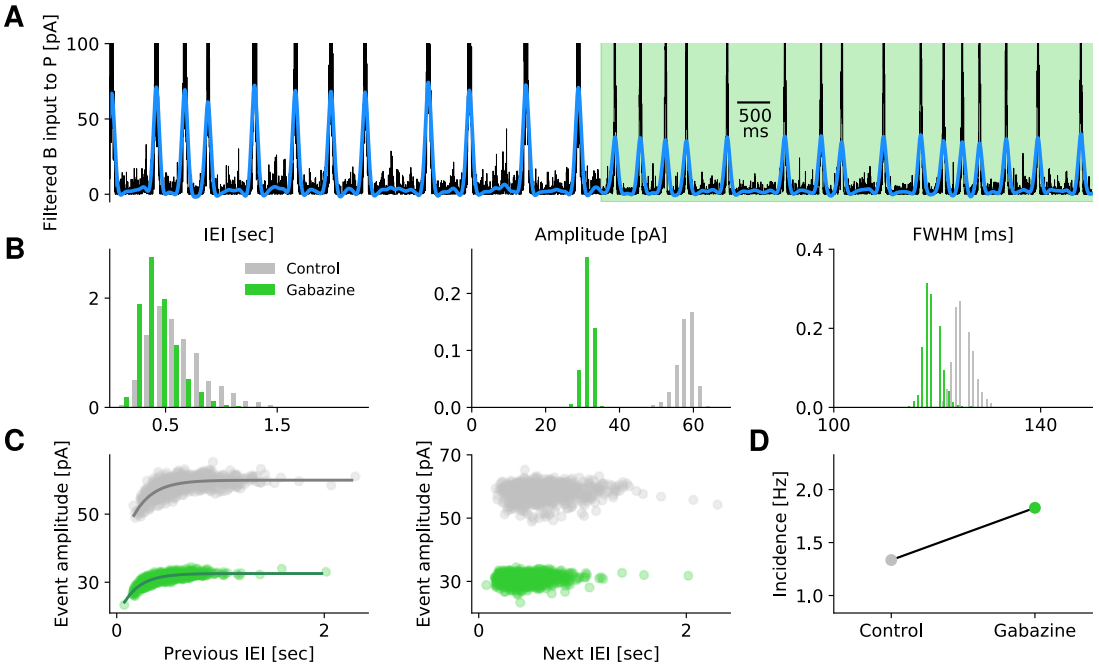

FIGURE 2.9: **Strongly reducing all inhibitory conductances (−50%) increases SWR incidence.** All inhibitory conductances have been reduced by 50% to mimic the effect of gabazine. The plot is organized as Fig. 2.8. Note the strong increase of SWR incidence under the gabazine condition (panel D).

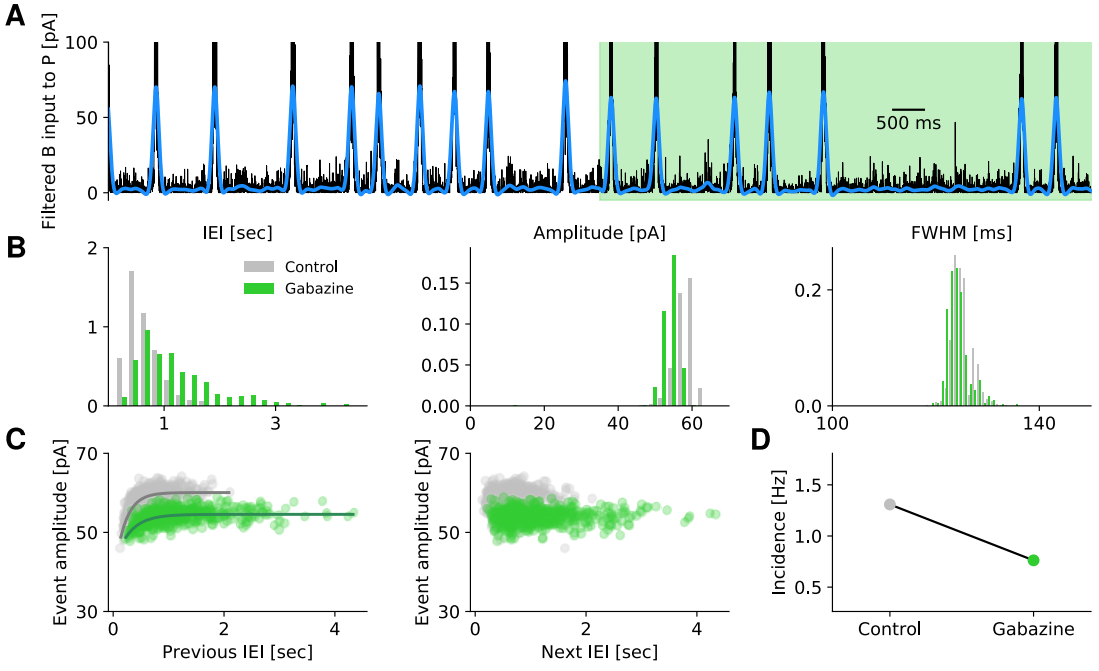

FIGURE 2.10: **Selectively reducing the disinhibitory conductance (g^{AB}) decreases SWR incidence.** The inhibitory conductances of the connection $B \to A$ have been reduced by 10% to mimic the selective effect of low doses of gabazine. All other connections are unaffected. The plot is organized as Fig. 2.8. Note the strong decrease of SWR incidence under the selective gabazine condition (panel D).

not when all connections are affected in the same proportion. The results on the non-selective gabazine application are in line with simulation results in a 2d network for sequence replay presented in Chenkov [2017].

Currently, there is no consensus on the effect of gabazine on SW amplitude, as both a decrease and an increase of SW amplitude following gabazine application have been reported. The observed difference might be due to the locality and larger doses of gabazine application in Schlingloff et al. [2014] compared to Chenkov [2017]. The model predicts that the SW amplitude is reduced in the gabazine condition, even though the firing rates of the populations vary depending on the paradigm of gabazine application. Future insights on the contribution of neurons' activity to the LFP could help clarifying whether the effect of gabazine on the different components reflects the model's prediction.

Overall, the analysis presented in this section predicts that also in the biological system, the effect of gabazine (and, more in general, of induced changes in the strength of the connections) might strongly depend on the exact details of the application paradigm, e.g., drug doses and location.

2.6 The spiking network as a perturbed bistable system

In the previous sections I have shown that the spiking network framework is suited to reproduce many experimental features, and to make testable predictions about the SWR generation mechanism. Nevertheless, a deeper understanding of the state of the network and the dependence on its parameters is still lacking. The large number of parameters involved prevents an easy understanding of how networks like the one introduced in Sec. 2.2 can be constructed.

The simulations of the spiking network presented in Fig. 2.2 show that the firing rate dynamics of the three populations evolves over time due to their interplay with the synaptic depression mechanism. Specifically, I have shown that the non-SWR state, which is characterized by low activity of P and B cells and by active A cells, is briefly perturbed by the emergence of a SWR event, which is characterized by high activity of P and B cells, and almost silent A cells. The system goes back to the initial (non-SWR) state due to the activation of the short-term depression mechanism at the $B \to A$ synapses. This behavior suggests that if the depression was ineffective in terminating an event, the system could be lingering in the SWR state for longer times.

This intuition is confirmed in Fig. 2.11, where the synaptic efficacy variables e_{ij} (regulating the depression strength of each synapse, see Eq. (2.3) in Sec. 2.8.2 for details) are artificially clamped at an intermediate value of $e_{ij} = 0.5$ (for all $B \to A$ synapses, so that the average synaptic efficacy is $e_{IJ} = 0.5$). In this scenario, the SWR state (labeled by B in Fig. 2.11) is initiated upon a depolarizing current injection to the P cells, and the SWR state is sustained until a hyperpolarizing current is applied. Thus, for fixed, intermediate values of the synaptic efficacy e_{IJ}, the system displays bistability: two stable states (A and B in Fig. 2.11) coexist in the network. Each of them is dominated by one active population of interneurons: A cells in the non-SWR periods, and B cells

in SWR periods. The system can switch between the two stable states due to external stimulation (in this example, current injection to P cells). Additionally, the transition can be induced by varying the synaptic efficacy (without current injection). Increasing the synaptic efficacy to large values (from $e_{IJ} = 0.5$ to $e_{IJ} = 0.8$ in Fig. 2.11) can result in a jump to a (long-lasting) SWR state: the intrinsic network fluctuations can drive the system to the SWR state. Note the delay ($\sim$25 ms) between the clamping of the synaptic efficacy to $e_{IJ} = 0.8$ and the start of the SWR state; the length of the delay decreases with increasing synaptic efficacy value and strength of fluctuations (not shown). In the extreme situation of a noiseless system, the system is not expected to switch to the SWR state. On the other hand, when the synaptic efficacy is decreased to $e_{IJ} = 0.2$, the non-SWR state is reset, and the system rests in this state in absence of further network modifications.

The two coexisting stable states are characterized by opposite activation of the interneurons: the competition between them is the mechanism that effectively regulates the alternation between SWR and non-SWR states. Thus, the network shown in Fig. 2.2 is in essence a perturbed bistable system, where the time evolution of the synaptic efficacy variable makes the two states weakly unstable and thus allows the network to switch between them. The existence of bistability for intermediate values of synaptic efficacy could already be guessed from Fig. 2.2, where the value $e_{IJ} = 0.5$ is associated to both the SWR state (descending part of the synaptic efficacy trace) and the non-SWR state (ascending part of the trace). I employ the idea of alternation between stable states to construct the spiking network, as explained in detail in Sec. 2.8.3.

2.6.1 On the stationary profile of the spiking activity

To better understand how the properties of the network change between SWR and non-SWR states (with clamped synaptic efficacy $e_{IJ} = 0.5$), I study in this section the f-I curves of the neurons in each population. I will make use of this distinction between states in Chapter 3, where the single neurons f-I curves will be used to describe the population dynamics.

In each of the states, the neurons receive synaptic input that depends on the firing rate of all presynaptically connected neurons in the network. As the firing rates of the populations are drastically different in the two states, we expect the input levels to be also different in either state. To better visualize this effect, Fig. 2.11 shows the mean f-I curves for each population in each stable state (for $e_{IJ} = 0.5$, see Sec. 2.8.5 for details on the definition). The shaded areas describe the distribution of input currents arriving, on average, to a neuron of a given population in either state; indeed, we can see that they are quite different. For example, for B cells, the input in the non-SWR state is $I \approx 57 \pm 78$ pA (mean $\pm$ one standard deviation), and, in the SWR state, $I \approx 277 \pm 173$ pA. Furthermore, the different input levels characteristic of either state also affect the shape of the f-I curves. Indeed, it is known that the f-I curve of a neuron receiving noisy inputs from other cells in the network can deviate quite strongly from the

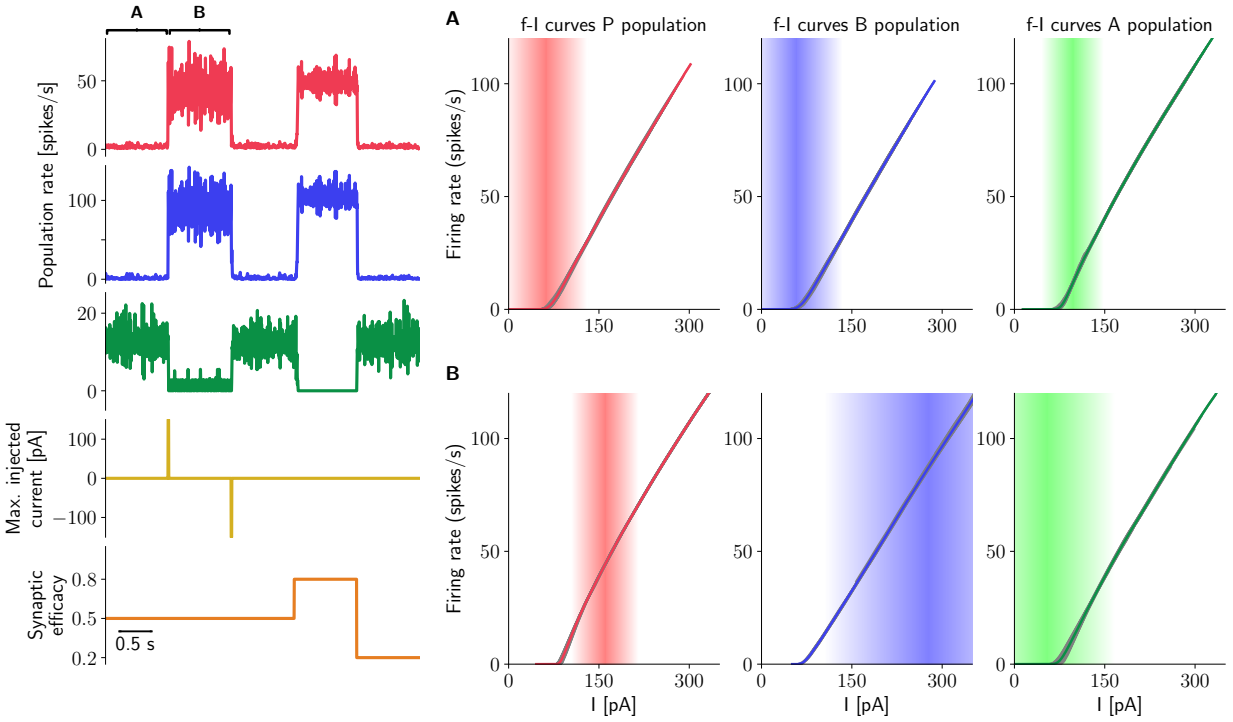

FIGURE 2.11: **The spiking network is a perturbed bistable system.** Left: Simulation results as synaptic efficacy is clamped at different values. For $e_{IJ} = 0.5$ (average value of synaptic efficacies of synapses $B \rightarrow A$), two stable states exist. Depolarizing current injection to P cells can switch the system from the non-SWR state to the SWR state. Population rates for the non-SWR state (state A) are: $P = 1.94$ spikes/s, $B = 1.32$ spikes/s, $A = 12.56$ spikes/s, and, during the SWR state (state B): $P = 43.60$ spikes/s, $B = 91.87$ spikes/s, $A = 1.12$ spikes/s. For $e_{IJ} = 0.8$, the network jumps to the SWR state due to internal fluctuations. The non-SWR state is restored for $e_{IJ} = 0.2$. Right: f-I curves of the spiking network in states A and B. Each row shows the f-I curves for P, B, and A cells (from left to right) in each stationary state (A: non-SWR state, top; B: SWR state, bottom). Gray lines are the f-I curves of single cells driven by external currents of different intensities. The curves are shifted on the x-axis to account for current from incoming synaptic inputs (see Sec. 2.8.5). Mean curves are depicted as colored solid lines. The shaded areas depict the regions where most inputs arrive (mean input current $\pm$ one standard deviation).

f-I curve of the neuron considered in isolation [Fellous et al., 2003; Gerstner et al., 2014; Shomali et al., 2018]. In general, the slope of the linear part (corresponding to large input levels) is preserved, whereas the close-to-threshold behavior can be drastically affected. Indeed, in our setup, the pairs of population f-I curves are parallel in their linear part, but the behavior close to threshold changes in the two states.

As only two stable states are present in the network, the set of two mean f-I curves defines the complete stationary firing behavior of each neuronal population in the network. I will take advantage of this feature in Chapter 3, where I aim at using the concept of network bistability to understand how to build networks that fulfill the desired properties.

2.7 Discussion

In this chapter I have shown that a spiking network consisting of pyramidal cells and two types of interneurons (PV^+ basket cells and a class of anti-SWR cells), equipped with a synaptic depression mechanism at the synapses connecting PV^+ basket cells to anti-SWR cells, is able to generate SWRs and to reproduce multiple features of experimentally recorded SWRs. SWRs can emerge spontaneously in the network or be triggered by cell stimulation (activation of pyramidal and PV^+ basket cells, inactivation of anti-SWR cells). The crucial mechanism underlying this behavior is based on the disinhibition of pyramidal cells via suppression of anti-SWR cells by active PV^+ basket cells.

The main prediction of the model is the existence of a population of interneurons, the anti-SWR cells, which are tonically active in non-SWR states and stop firing during SWR events. I also predict that inactivating these cells is sufficient to trigger a SWR event. In Chapter 4, I show that the existence of this class of cells seems to be supported by experimental results. Additionally, the analysis of the timing between successive SWRs suggests that the synaptic depression mechanism could play an important role in regulating the incidence and occurrence of SWRs. This hypothesis could be tested in experiments by interfering with the recovery of the PV^+BCs-dependent depression mechanism and studying whether this manipulation affects SWR timing.

Several assumptions have been made to define the spiking model. First, I decided to model each cell as a leaky integrate-and-fire unit. This approach has the advantage of keeping the number of variables associated with the spiking generation mechanism relatively low (compared to more complex models including the ion channel dynamics, see e.g. [Hodgkin and Huxley, 1952]). The main drawback of this choice is that the shape of the action potential is not explicitly modeled. However, I believe that this simplification does not affect the main features of the model. In fact, as the analysis presented in Chapter 3 suggests that single neurons play a minor role compared to the dynamics of the whole population, I expect the contribution of the shape of single action potentials to be negligible. To model the interaction between cells, I have assumed that each incoming spike causes an increase in the conductance of a given neuron. If multiple spikes arrive in a relatively short time window, their contribution is summed up and they can elicit a spike in the target neuron. Compared to the simpler alternative of current-based synapses (each incoming spike contributes directly to the input current of a neuron), the conductance-based approach is closer to the experimental scenario and allows to better link the model parameters with observable biological quantities. In Chapter 3, I explore the impact of this choice on the analysis of the network behavior and argue that the choice of current- vs. conductance-based synapses largely depends on the scope of the model (analytically tractable model or biological realism).

A standard way to connect neurons of a spiking network consists in assuming that the connections among two populations are random, i.e., no structure is present. This simplification does not take into account distance-dependent connection probabilities

and connectivity motifs that have been observed in biological networks [Song et al., 2005; Perin et al., 2011; Rieubland et al., 2014; Guzman et al., 2016; English et al., 2017]. Together with the choice of standard neuronal properties for each neuron in a given population, this assumption guarantees that the spiking network is homogeneous, i.e., the activity of every neuron is representative of its population. In Chapter 3, I make use of this property to approximate the spiking network with a model where only the population activities, and not the behavior of each single cell, is modeled.

As a possible extension of the model, inhomogeneous populations could be considered. For example, it would be interesting to test the impact of a larger cell-to-cell variability (e.g., in spiking thresholds, reset potentials, etc.), and to include different subpopulations of P, B, and A cells (see, e.g., the recent finding by Hunt et al. [2018] and Chapter 5). I expect that the main features of the network are preserved, assuming that the variability follows a Gaussian distribution around the value used in the default simulation (see Table 2.1). Additionally, heterogeneity at the level of the connections could be introduced. This choice could help explaining a number of experimental features that cannot be replicated in the current, homogeneous model. First, it has been shown by Bazelot et al. [2016] that SWR events can be triggered by driving a single pyramidal cell to spike (even a single action potential can be sufficient). This result cannot be replicated in the model, where the activation of at least ∼20-30 pyramidal cells is needed to elicit a SWR event. This limitation of the model is due to the large size of the pyramidal cell population, i.e., each neuron contributes little to the depolarization of connected cells. However, if cells were connected in a non-random fashion, it could be possible for a pyramidal cell with a large number of postsynaptic targets to be the initiator of a SWR. Second, inhomogeneous networks could also replicate the experimental finding that only up to 50% of pyramidal cells are involved in a single SWR event [Ylinen et al., 1995; Ellender et al., 2010]. Depending on the differential embedding of cells in the network, fewer cells could be recruited in each event. Interestingly, the aspect of cell participation is linked to the evidence of sequence replay during experimentally observed SWRs *in vivo* (see Sec. 1.1.3.3). If cells were organized in small-size clusters (cell assemblies) of strongly connected cells coding for a specific memory, only the assemblies related to the currently reactivated memory would be active in a given SWR event, thus lowering the proportion of recruited cells during a single event.

We have already discussed the importance of the connection $B \rightarrow A$ for the generation of both spontaneous and induced SWRs. However, as anti-SWR cells are yet to be identified, the existence of this connection is still unknown. It is known that PV^+ basket cells contact different classes of inhibitory neurons [Sik et al., 1995; Cobb et al., 1997; Kohus et al., 2016; Walker et al., 2016], but it would be important to test if a monosynaptic connection to anti-SWR cells exists, and if it shows plastic activity modulation. In the model, short-term synaptic depression at these synapses is crucial to terminate SWRs. This choice was inspired by Kohus et al. [2016], who showed that SWR occurrence correlates with a depression mechanism from PV^+ basket cells to pyramidal cells. I predict that a similar dynamics is present at the $B \rightarrow A$ synapses, but this remains to be tested.

In Sec. 2.5.1, I illustrated alternative plasticity mechanisms, which could support, or even replace, the depression at the $B \to A$ synapses. Alternatively, the depression mechanism could be replaced by spike frequency adaptation in the B cells [Kneisler and Dingledine, 1995; Povysheva et al., 2013; Ha and Cheong, 2017], even though it is currently unclear whether PV^+ basket cells have this property.

Another aspect that needs attention is the importance of the relative strengths of converging pathways (Sec. 2.3.1). This characteristic is analyzed in more detail in Chapter 3, and it would be interesting to test in experiments how SWRs dynamics change when the relative strength of the pathways is altered. I predict that the network is still capable of generating SWRs, but possibly only in a subset of experimental conditions (e.g. activation of P cells, but not of B cells).

At a first glance, it might seem that the conditions on pathway strength strongly constrain the set of parameters that yield networks showing the desired behavior. However, it is important to mention that the spiking network presented in this chapter did not require fine tuning of parameters. Indeed, it will become clear in Chapter 3 that the desired network configuration can be obtained for many parameter combinations.

How could the connections be tuned so that all desired properties of the network are fulfilled? As discussed in Sec. 2.2, the building blocks of the model are the two subnetworks P-A and P-B, with the reciprocal interneuronal connections allowing to switch between states dominated by one of the two subnetworks (the other interneurons are almost inactive, also see Sec. 2.8.3). An interesting hypothesis is that the two subnetworks could be tuned using some self-organizing principle. A possible scenario could feature inhibitory spike-timing-dependent plasticity (STDP, see Woodin et al. [2003]; Vogels et al. [2011]; Luz and Shamir [2012]; Vogels et al. [2013]) to adjust the inhibitory-to-excitatory connections in each subnetwork. As presented in Vogels et al. [2011], inhibitory STDP adjusts the values of the conductance updates of the inhibitory-to-excitatory connections to bring the firing rate of the pyramidal cells close to a given target value. In our scenario, we can use experimentally known values of P cells activity in the SWR and non-SWR states as a target. By adjusting the strength of inhibition, STDP stabilizes the network dynamics, balancing the excitatory and inhibitory inputs that a neuron is receiving. The authors claim that for a large parameter range, the stabilized population activity is characterized by irregular and asynchronous firing (AI state [1]), a state which is thought to well describe cortical dynamics [Softky and Koch, 1993; Csicsvari et al., 1999b; Okun and Lampl, 2008; Bhatia et al., 2019]. In the setting of my model, this approach can successfully be used to tune the networks to represent the dynamics of the non-SWR and SWR states; however, it has two main drawbacks. First, in the SWR state the firing of P cells is large ($P \approx 43$ spikes/s, see Fig. 2.11) and the inhibitory STDP cannot (in my hands), bring the network to an AI state. Furthermore, STDP disrupts the randomness of the connections, a useful property of the spiking network that will be used in Chapter 3 to define an analytically treatable reduction of the model. Overall,

[1] This means that single cells fire irregularly, and that the firing of multiple cells is not synchronous.

even though the STDP approach is not entirely satisfying, it can be considered as a first step towards the modeling of self-organizing networks displaying SWRs.

Spiking models can display spontaneous SWRs. The emergence of spontaneous events depends on the interplay between the reciprocal connections among interneurons and the parameters of the synaptic depression mechanism (see Sec. 2.8.3). More effort should be devoted to understanding how each parameter affects the existence of spontaneous SWRs. In particular, realistic values of SWR duration and incidence, and the emergence of the correlation structure described in Fig. 2.4, currently require manual tuning of the aforementioned parameters. The development of a more theoretical approach would be needed to better understand under which conditions the network displays spontaneous events. Specifically, the analysis should yield a quantification of the network-driven noise in the system, because the fluctuations are crucial to initiate a SWR and they ultimately define how much the network deviates from a bistable configuration (where non-SWR and SWR states coexist). A good starting point for the analysis could be the theory of balanced networks [van Vreeswijk and Sompolinsky, 1996; Van Vreeswijk and Sompolinsky, 1998; Brunel, 2000a; Kumar et al., 2008; Renart et al., 2010], even though the assumptions might be partially violated in the SWR state (where the populations show synchronous events, the ripples). Note that the three-population structure of the network makes the analytical work more challenging than the standard 2d excitation-inhibition scenario. This approach could open an exciting direction of research, to study how the noise-driven behavior of the network is influenced by neuronal properties.

Finally, I have shown in Sec. 2.5.2 that the spiking network is suited for investigating the effect of pharmacological modulation of spontaneous SWR features. I have focused on gabazine, a $GABA_A$ receptors antagonist, due to its surprising effect of decreasing SWR incidence in hippocampal slices. This effect is surprising because down-regulating inhibitory conductances is expected to increase the excitation in the network, leading to more (and not fewer, as observed in Nimmrich et al. [2005]; Chenkov [2017]) SWR events. In the simulations of the spiking model, we observed that a selective reduction of the strength of the connection $B \to A$ strongly reduces the SWR incidence. However, the results are less clear when all inhibitory conductances are affected equally. Thus, it would be interesting to test in slice experiments whether gabazine affects all inhibitory conductances homogeneously, or whether differences might arise depending on the synapse type. The analysis has shown a nonlinear effect of gabazine application when multiple connections are affected, a feature which might, at least in part, explain the controversial results reported in the literature (on, e.g., the effect on SW amplitude, see Schlingloff et al. [2014]; Chenkov [2017]). To shed light on the effect of gabazine modulation, it could be helpful to resort to a theoretical approach, with the aim of predicting the effect of changing connection strengths on SWR features. Some steps in this direction were done by Chenkov [2017], for a 2d hippocampal network that models hippocampal sequence

replay. The approach could be extended to the analysis of the mean-field model I present in Chapter 3.

Even though gabazine has been widely studied in relation to SWRs, there are other drugs whose modulation has been reported to influence SWR dynamics. An extension of the work presented here could include the modeling of the effects of thiopental and diazepam on SWRs [Papatheodoropoulos et al., 2007; Koniaris et al., 2011]. Donoso et al. [2018] focused on the effect of these and other drugs on a model of ripple oscillations, but it would be interesting to also study the impact on the sharp wave component. Furthermore, experimental and theoretical efforts have been directed to understand the role of $GABA_B$ receptors (the second main type of GABA receptor in the brain) in modulating SWRs. In contrast with reports suggesting that blocking $GABA_B$ receptors does not alter SWR properties [Hollnagel et al., 2014; Hofer et al., 2015], a recent analysis [Chenkov, 2017] suggests that the application of a $GABA_B$ antagonist (SCH50,911) increases SWR incidence. Additionally, the application of a $GABA_B$ agonist (baclofen) has been reported to decrease SWR incidence [Hollnagel et al., 2014]. As thoroughly reviewed in Chenkov [2017], an understanding of the $GABA_B$ receptor-mediated mechanisms influencing SWR dynamics is still lacking.

2.8 Methods

2.8.1 Neuron model

Neurons are modeled as conductance-based leaky integrate-and-fire units [Stein, 1967]. The subthreshold membrane potential $V_i(t)$ of cell i obeys

$$C\frac{\mathrm{d}V_i}{\mathrm{d}t} = g_L(V_{rest}-V_i)-[g_i^P(t)(V_i-E_{rev}^P)+g_i^A(t)(V_i-E_{rev}^A)+g_i^B(t)(V_i-E_{rev}^B)]+I_{ext} \quad (2.1)$$

where $C = 200$ pF is the membrane capacitance and $g_L = 10$ nS is the leak conductance, resulting in a membrane time constant $\tau = 20$ ms. $V_{rest} = -60$ mV is the resting membrane potential, $E_{rev}^P = 0$ mV, $E_{rev}^B = -70$ mV, and $E_{rev}^A = -70$ mV are the reversal potentials of excitation and inhibition (of B and A cells, respectively), and $I_{ext} = I_{BG} + I_i$ is the sum of external currents. To elicit activity in the network, a constant background current $I_{BG} = 200$ pA is injected to all neurons. Only if explicitly mentioned, neurons receive additional time-dependent currents I_i. Every time a neuron's membrane potential reaches the threshold $V_{thr} = -50$ mV, a spike is emitted and V_i is reset to the reset potential (for simplicity, it equals V_{rest}), where it is clamped for a refractory period of length $t_{refr}^I = 1$ ms, $I \in \{P, B, A\}$.

The outgoing synapses from pyramidal cells are modeled as fast AMPA-type synapses, and the synapses originating from B or A cells are modeled as $GABA_A$-type synapses (see e.g. [Ellender et al., 2010] for motivation). The variables $g_i^P(t)$, $g_i^B(t)$, and $g_i^A(t)$ describe the total synaptic conductances resulting from incoming synaptic inputs to neuron i. To simplify the notation, the time dependence is dropped. The conductance dynamics are

described by

$$\begin{cases} \dfrac{\mathrm{d}g_i^P}{\mathrm{d}t} = -\dfrac{g_i^P}{\tau_{syn}^P} + \sum_{f,j} \delta(t - t_j^{(f)} - \tau_{IP}) g_{ij}^{IP} \\ \dfrac{\mathrm{d}g_i^B}{\mathrm{d}t} = -\dfrac{g_i^B}{\tau_{syn}^B} + \sum_{f,j} \delta(t - t_j^{(f)} - \tau_{IB}) g_{ij}^{IB} \\ \dfrac{\mathrm{d}g_i^A}{\mathrm{d}t} = -\dfrac{g_i^A}{\tau_{syn}^A} + \sum_{f,j} \delta(t - t_j^{(f)} - \tau_{IA}) g_{ij}^{IA}, \qquad I \in \{P, B, A\} \end{cases} \tag{2.2}$$

where $\delta(t - t_j^{(f)} - \tau_{IJ})$ is the contribution of the f-th incoming spike (from neuron j at time $t_j^{(f)}$); δ is the Dirac delta function. The quantities g_{ij}^{IP}, g_{ij}^{IB}, and g_{ij}^{IA} describe the unitary conductance updates resulting from a single spike. For example, g_{ij}^{IP} is the conductance update by a P presynaptic neuron j connected to neuron i in population $I \in \{P, B, A\}$ (i.e., these values depend on the synapse type). There is a delay between a presynaptic spike and the postsynaptic response onset defined as $\tau_{IJ} = 1$ ms for all synapse types. The conductances decay exponentially with time constants $\tau_{syn}^P = 2$ ms, $\tau_{syn}^B = 1.5$ ms, and $\tau_{syn}^A = 4$ ms [Geiger et al., 1995; Bartos et al., 2002; Taxidis et al., 2012]. For simplicity, time constants only depend on the pre-, but not the postsynaptic type. The values of the unitary conductance updates range from 0.2-8 nS and are listed in Table 2.2.

2.8.2 Network model

I model a network comprising $N_P = 8200$ pyramidal cells (P), $N_B = 135$ PV$^+$ basket cells (PV$^+$BCs, B in the model), and $N_A = 50$ anti-SWR cells (A) cells. These numbers are chosen to mimic the number of P and B cells present in CA3 in a 400 μm-thick rat slice. Indeed, it has been estimated that the entire rat hippocampus contains 204.700 pyramidal cells and 25.300 interneurons in the CA3 region [Bezaire and Soltesz, 2013]. Given that a 400 μm-thick slice represent approximately 4% of the volume of the rat hippocampus, I estimate that around 8200 pyramidal cells are present in a slice. In CA1, PV$^+$ basket cells are thought to account for approximately 14% of all interneurons. As we do not have a closer estimate for CA3, I assume the same holds in CA3, yielding approximately 135 PV$^+$ basket cells in a CA3 slice [2]. Given that the identity of anti-SWR cells is unknown, no such data are available for these cells; I decided to include 50 anti-SWR cells in the network.

Neurons are randomly connected with connection probability p^{IJ} for connection $J \to I$. In contrast to the dominant view of CA3 as a strongly recurrent region, it was recently shown that CA3 pyramidal cells are, at least *in vitro*, only sparsely connected [Guzman et al., 2016]. I thus choose $p^{PP} = 0.01$. Recurrent connectivity among PV$^+$ basket cells

[2] The numbers in Bezaire and Soltesz [2013] are not precise: the authors assume that 11% of all neurons are interneurons, which yields 22.517 interneurons, but they estimate the number of CA3 interneurons as 25.300. The estimate of 135 PV$^+$BCs is calculated as the 14% of 24.100 interneurons in a volume of 4%.

is usually estimated to be around 20% (in rat CA1 [Sik et al., 1995; Donoso, 2016] and in mouse CA3 [Schlingloff et al., 2014]). A recent study [Kohus et al., 2016] suggested that connectivity could be as high as 66% (in mouse CA3, *in vitro*); nevertheless, I consider the conservative estimate of 20% and thus set $p^{BB} = 0.2$. There is a large body of work that studies the bidirectional connectivity between pyramidal cells and interneurons; however, only few studies are specific for PV^+ (possibly basket) cells, and none of these addresses CA3. Mouse CA1 studies [Lee et al., 2014] suggest that the connectivity from PV^+BCs to pyramidal cells could be in the range of 45-50%, and the one from pyramidal cells to PV^+BCs approximately 16-48%. I choose $p^{BP} = 0.2$ and $p^{PB} = 0.5$. The connectivity from and to anti-SWR cells cannot be constrained by experiments until more is known about the identity of these cells. The connectivity values I choose are: $p^{AP} = 0.01$, $p^{PA} = 0.6$, $p^{AA} = 0.6$, $p^{AB} = 0.2$, $p^{BA} = 0.6$. The values p^{PA}, p^{BA}, and p^{AA} ensure that each neuron in the postsynaptic population receives at least 5 synapses from A cells. Indeed, it has been shown that the number of connections in spiking networks cannot be reduced to too few synapses [Chenkov et al., 2017].

The network is constructed such that in the non-SWR state, the P and A populations are in an asynchronous irregular (AI) regime – a state that is thought to reflect the state of the hippocampus at rest [Bhatia et al., 2019]. In this state, population firing rates are tuned to have P cells firing at around 2 spikes/s, A cells at around 12 spikes/s, and B cells to be almost inactive, with average firing rates lower than 2 spikes/s. The SWR state is dominated by a strongly active P-B subnetwork, where P cells fire at 43 spikes/s, B cells fire at around 90 spikes/s and A cells are almost inactive, with average firing rates lower than 2 spikes/s.

A short-term synaptic depression mechanism is present at the $B \to A$ connections, which modulates the strength of the unitary synaptic conductance updates. The synaptic updates g_{ij}^{AB} from neuron j in population B to neuron i in population A are scaled by a factor e_{ij} describing the synaptic efficacy, which evolves over time as

$$\frac{de_{ij}}{dt} = \frac{1 - e_{ij}}{\tau_D} - \sum_f \delta(t - t_j^{(f)}) e_{ij} \eta_D. \tag{2.3}$$

Every time a cell j in population B spikes, the g_i^B conductance for the connected postsynaptic cells i is updated by the product $e_{ij} g_{ij}^{AB}$ (instead of only g_{ij}^{AB} as in the non-depressed case, see Eq. (2.2)), and the e_{ij} variables of all synapses starting from the spiking cell j are decreased by an amount $e_{ij}\eta_D$. Hence, higher activity (i.e. more spikes per second) of one cell in population B results, on average, in a lower efficacy of synaptic transmission to its connected cells in population A. To prevent the emergence of negative conductance updates, e_{ij} is restricted to the interval $[0, 1]$ through the dynamics described in Eq. (2.3). The depression mechanism, with values chosen as $\eta_D = 0.18$ and $\tau_D = 250$ ms, is responsible for the termination of a SWR event and, more in general, for driving the system back to the non-SWR state. In the main text, the synaptic efficacy variable e_{IJ} defines the averaged value of e_{ij} across all $B \to A$ synapses.

If not explicitly mentioned, all other conductance updates g_{ij}^{IJ} are kept fixed. In specific cases (see Sec. 2.5.1), the $P \to A$ connection is considered to be plastic, with a short-term facilitation mechanism, and the $B \to P$ has a short-term depression mechanism analogous to the one described above. For the latter case, the g_{ij}^{PB} conductance is scaled by a synaptic efficacy variable described by Eq. (2.3). For the simulations shown in Fig. 2.5, I choose $\eta_D = 0.18$ and $\tau_D = 250$ ms, analogous to the $B \to A$ depression. All other parameters are unchanged.

The facilitation at the synapses $P \to A$ is modeled as follows: the conductance g_{ij}^{AP} is scaled by a factor $(1 + z_{ij})$ (synapse-dependent), where $z_{ij} \geq 0$ describes the effect of facilitation. In the case of no facilitation, $z_{ij} = 0$. The facilitation variables evolve over time as

$$\frac{dz_{ij}}{dt} = -\frac{z_{ij}}{\tau_F} + \sum_f \delta(t - t_j^{(f)})(z_{max} - z_{ij})\eta_F. \tag{2.4}$$

Every time a j cell in the P population spikes, the AMPA conductance g_i^P of a connected cell i (see Eq. (2.2)) is scaled by a factor $(1 + z_{ij})$, and the z_{ij} variables of all synapses whose presynaptic cell is j are increased by an amount $(z_{max} - z_{ij})\eta_F$. z_{max} is a constant value defining an upper bound for the increase in facilitation. When the system is in the non-SWR state, z_{ij} decays exponentially to the value $z_{\text{non-SWR}} = \frac{P_0 z_{max} \eta_F \tau_F}{1 + P_0 \eta_F \tau_F}$, where P_0 is the firing rate of P cells in the non-SWR state (around 2 spikes/s, see Fig. 2.11). To better compare the default network (with $B \to A$ depression) to the case where extra facilitation is added, I additionally normalize g_{ij}^{AP} by diving it by $(1 + z_{\text{non-SWR}})$. This assures that when the facilitation is active, but has already decayed to $z_{\text{non-SWR}}$, the $P \to A$ synapses have the same strength (i.e, the same conductances) as in the model with no facilitation. For the simulations shown in Fig. 2.6, I choose $\eta_F = 0.15$, $\tau_F = 250$ ms, $z_{max} = 1$. All the other parameters are as in the default model. For the simulation where $P \to A$ is the only plastic mechanism (Fig. 2.7), we need to adjust the parameters for the network to be in a regime where the non-SWR state is destabilized and events can start spontaneously with a large enough incidence (if the incidence is too low, we cannot observe correlation between IEI and event amplitude). To this end, I choose $g^{AB} = 4.5$, $g^{BA} = 5.5$, $\tau_F = 230$ ms, $\eta_F = 0.32$, $z_{max} = 1$, and do not normalize g^{AP} by its non-SWR state value ($z_{\text{non-SWR}}$). Note that if the synaptic efficacy is frozen in the network shown in Fig. 2.6, no spontaneous events occur because the facilitation effect is counterbalanced by a too strong $B \to A$ connection (i.e. g^{AB} too large), which keeps the system in the SWR state (simulation not shown).

2.8.3 Constructing the network

In this section, I present a procedure to construct the spiking network shown in Fig. 2.2. I start by fixing the number of cells and the connection probabilities of P, B, and A cells using the values introduced in Sec. 2.8.2.

To tune the values of the unitary conductance updates g_{ij}^{IJ}, $I, J \in \{P, B, A\}$ (see Sec. 2.8.1), I rely on the observation that the two groups of interneurons B and A should

be active at different stages. B cells should be almost inactive in non-SWR states, and increase their firing during SWRs, whereas A cells should be tonically active throughout the non-SWR state and stop firing during it. Thus, both the non-SWR and SWR states are dominated by a 2d subnetwork of active cells: the pyramidal cells, and one type of interneurons. On a first approximation, I consider the firing rate of the other interneuron type as being close to 0 spikes/s.

For this reason, I construct the network starting from the P-A subnetwork, in isolation. I assume that the unitary conductance updates g_{ij}^{IJ} are the same across each i, j combination (i.e., they only depend on the synapse type), and choose the values g_{ij}^{PP}, g_{ij}^{AP}, g_{ij}^{PA}, and g_{ij}^{AA} such that the neurons in both populations fire asynchronously and irregularly (AI regime), with mean firing rates $P \approx 2$ spikes/s and $A \approx 12$ spikes/s. These values have been chosen to be close to experimental values (see Sec. 1.2.3 and Chapter 4), but the exact choice of the target firing rate values does not impact the results presented here. While choosing the conductance update values, I enforce the conditions on the pathway strengths (see Sec. 2.3.1) by selecting a small g_{ij}^{AP} (requirements 1 and 2) and a large g_{ij}^{PA} (requirements 3 and 5). These requests are relatively easy to fulfill, because g_{ij}^{PA} is expected to be large for the inhibition to stabilize the P-A subnetwork, and, vice versa, B cells should not receive too much excitation [3]. I monitor the coefficient of variation (CV) and the standard deviation (SD) of the instantaneous population rates (Gaussian filter time constant is 3 ms) [Vogels et al., 2011] to assure that the neurons fire fairly irregularly (CV > 0.5) and asynchronously (SD < 1).

Then, I focus on the isolated P-B subnetwork and tune the conductance update values g_{ij}^{PB}, g_{ij}^{BP}, and g_{ij}^{BB} such that P cells fire close to the frequency of $P \approx 43$ spikes/s, and B cells close to the frequency of $B \approx 90$ spikes/s. I use the value of g_{ij}^{PP} defined in the subnetwork P-A. Also in this case, firing rates have been chosen to be close to experiments (see Sec. 1.2.3), but other choices are also possible. Nevertheless, the firing rates of P, B, and A cells should be sufficiently different in the SWR and non-SWR states (at least $\sim$5 spikes/s difference) for the system to jump between clearly distinguishable states. Requirements on the strengths of pathways (see Sec. 2.3.1) are enforced by selecting a large g_{ij}^{BP} (requirements 1 and 2) and a small g_{ij}^{PB} (requirements 3 and 5). In fact, the choice of g_{ij}^{PB} is a compromise between these requirements and the fact that the connection $B \to P$ should be strong enough for the interneurons to control the spiking of P cells. As a result, it is difficult to obtain a network in an AI state: the units are firing regularly (CV < 0.1) and in synchrony (SD > 1). The synchronicity of unit firing is clearly visible in the power spectrum (the peak firing frequency is 135 Hz) and results in the ripple-like oscillations shown in Fig. 2.2.

Up to this point, I have built two subnetworks that display clearly distinguishable states of stable firing of P and A, and P and B cells, respectively. I now wish to connect the two subnetworks by defining the reciprocal connections between the interneurons. First, I add A cells to the P-B subnetwork, with connections $A \to A$ and $P \to A$ (from the P-A subnetwork simulations), and define a new connection $B \to A$ such that A cells

[3] For a more formal way of enforcing the requirements, see Chapter 3.

fire less than 2 spikes / s (i.e., are almost inactive in this state). The 'disinhibitory' connection $B \to A$ is expected to be large to control the firing of A cells and to comply with requirements 1 and 3 (see Sec. 2.3.1). This scenario is constructed to represent the SWR state, where I assume that the missing connections $A \to P$ and $A \to B$ play a negligible role because A cells are almost inactive.

Then, I simulate a network with all connections defined in the previous steps, and a new connection $A \to B$, and choose g_{ij}^{BA} such that B cells fire less than 2 spikes / s. To enforce the requirements on the pathway strengths (see Sec. 2.3.1), however, the connection $A \to B$ should not be too strong (requirements 2 and 5). This scenario corresponds to the non-SWR state.

The enforcement of the requirements presented in Sec. 2.3.1 guarantees that, upon cell stimulation, the firing rates of all populations change as requested. Note that I do not explicitly enforce the requirements 4 and 6. These conditions demand that at least one of two pathways ($B \to A$ and $B \to P \to A$, and $A \to B$ and $A \to P \to B$, respectively) is strong enough for the current injection to elicit the desired response, but these requisites are already included in the requirements 1, 2, 3, and 5.

The full network constructed with this procedure has two embedded stable states: one dominated by the P-A subnetwork (non-SWR state), and one dominated by the P-B subnetwork (SWR state). Thus, there is an intrinsic bistability structure in the network: external mechanisms (e.g. current injection) are needed to switch between the two states. The conductance update values g_{ij}^{AB} and g_{ij}^{BA} regulate the stability of the two states. They are chosen to be large enough to inhibit the inactive interneuron type in each state, but should not be too large, to guarantee that both states are stable. For example, even when initialized to be in the non-SWR state, a network with too strong $B \to A$ connection would spontaneously jump to the SWR state. This is because the low activity of the B cells is amplified by the strong $B \to A$ connection and suffice to inhibit the A cells.

In the last step, I add the synaptic depression mechanism at the $B \to A$ synapses (with dynamics described by Eq. (2.3)), which is responsible for the termination of the SWR state.

To generate a network exhibiting spontaneous SWRs, I destabilize the non-SWR state by modifying the reciprocal connections between the two interneurons (in particular, their conductance update values g_{ij}^{AB} and g_{ij}^{BA}). Indeed, decreasing the strength of the $A \to B$ connection decreases the inhibition of the B cells, and increasing the strength of the $B \to A$ connection makes spontaneous jumps to the SWR state easier, because it promotes the inhibition of A cells by B cells. Together with the choice of the depression parameters τ_D and η_D, these changes allow fluctuations in the activity of B cells to start a SWR event. Note that the definition of a network with spontaneous events with the right correlation structure (see Fig. 2.4) is complicated by the interplay of the different parameters (g_{ij}^{AB}, g_{ij}^{BA}, τ_D, η_D) and requires manual tuning. I resort to a grid search procedure to select the parameters that yield spontaneous SWR with incidence and

correlation structure close to experiments. Finally, similarly to the scenario of evoked SWRs, the depression mechanism is responsible for event termination. The connectivity parameters used to simulate the network presented in Fig. 2.2 are listed in Tables 2.2 and 2.3.

2.8.4 Simulation analysis

All simulations are performed in Brian [Goodman and Brette, 2009], and data analysis is performed in Python (`www.python.org`). Population firing rates are computed by averaging the instantaneous firing rates, averaged across neurons, with a Gaussian smoothing window with standard deviation $w = 3$ ms. I use the modulation of population firing rates as a signature of a SWR event: an increase of P cells firing to $P \approx 43$ spikes/s, an increase of B cells firing to $B \approx 90$ spikes/s, and a decrease of A cells to values lower than 2 spikes/s mark the start of a SWR event. All conditions have to be simultaneously fulfilled for a SWR event to be detected. To trigger a SWR event, I randomly select 60% of the cells in a given population and stimulate them with currents uniformly distributed between $I = 0$ pA and a maximal value $I_P = 400$ pA, $I_B = 500$ pA, or $I_A = -500$ pA for $T = 10$ ms. The short simulation times are comparable with the duration of optogenetic stimulation used in Schlingloff et al. [2014]; Kohus et al. [2016]. Stimulation results are hardly affected by differences in the stimulation parameters, as far as the stimulation paradigm is sufficient to initiate a SWR event. In all simulations shown in this chapter, the variables of Eqs. (2.1) and (2.2) (V_i, g_i^P, g_i^B, g_i^A) are initialized for the system to be in the non-SWR state.

To define the LFP in *stratum pyramidale*, I assume that the main contribution to the field is provided by perisomatically-targeting interneurons [Beyeler et al., 2013; Schönberger et al., 2014], namely PV$^+$BCs, targeting the cell bodies of pyramidal cells. A main criticism to this approach is that the cell morphology and non-perisomatic (e.g. dendritic) inputs might also contribute to shape the LFP [Einevoll et al., 2013; Chizhov et al., 2015]. However, as a detailed description of the LFP is beyond the scope of the simplified point neuron scenario considered here, I resort to this simple approach to define an approximated LFP trace. Note that I implicitly assume that anti-SWR cells do not contribute to the LFP. This assumption could hold if anti-SWR cells target pyramidal cells at the distal dendrites, so that their contribution at the pyramidal cells somata can possibly be neglected. Notably, Fig. 2.12 shows that there are almost no anti-SWR-cells-related currents impinging onto pyramidal cells during SWRs (our events of interest), because most A cells are inactive. Thus, I entirely focus on the contribution from PV$^+$BCs to define the LFP.

In summary, I define the LFP as a filtered version of the synaptic input current from B to P cells, averaged over all $B \to P$ synapses (or a subset of synapses, where explicitly stated). To obtain the sharp wave and ripple components of a SWR event, I filter the sign-reversed mean B input current to P cells in two different frequency

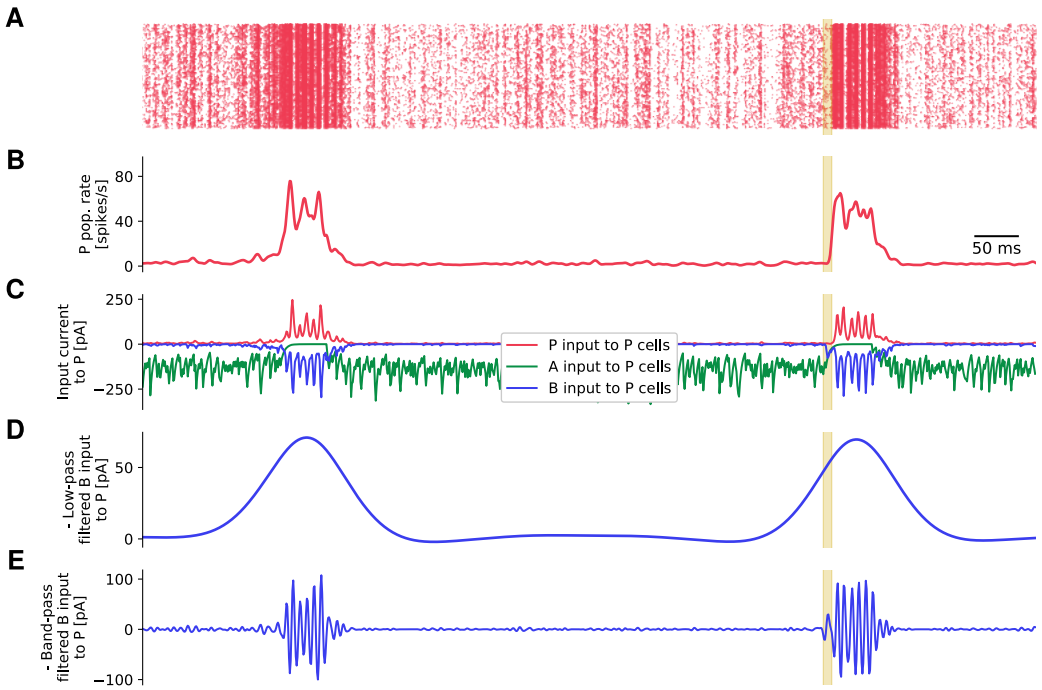

FIGURE 2.12: **Definition of approximated LFP. A)** Raster plot showing a spontaneous and an evoked SWR event. **B)** Average population firing rate of P cells. **C)** Input current from P (red), B (blue), and A (green) cells impinging onto pyramidal cells, averaged across all neurons. The averaged input current from B to P cells is sign reversed and used as an approximation of the LFP. **D)** LFP signal shown is low-pass filtered up to 5 Hz to extract the sharp wave component. **E)** LFP is band-pass filtered in the 90-180 Hz range to extract the ripple component.

bands, using a Butterworth filter of order 2. The sharp wave component is obtained by low-pass filtering the signal up to 5 Hz. The cutoff frequency is chosen for the filter to cover the whole duration of a postsynaptic event. The ripple component is obtained by band-pass filtering the signal in the range 90-180 Hz, around the peak frequency of 135 Hz (the peak is computed in the power spectrum). For spontaneous and evoked SWR events, I define the following properties: inter-event-interval (IEI), amplitude, and full width at half maximum (FWHM). The *amplitude* is the peak value of the sharp wave (SW) component (i.e., the low-pass filtered LFP signal; peaks are detected using a script available at Duarte [2015]). To compute the *FWHM* I first define the mean baseline value as the mean across all events of the average value of the low-pass filtered signal in periods preceding a sharp wave by 200 to 100 ms. Then, I calculate the half maximum by finding the mean value of the event amplitude and the mean baseline value. I define the start of a sharp wave event as the time of the sharp wave signal at half maximum preceding the peak, and the event end as the time of the sharp wave signal at half maximum following the peak. The *IEI* is defined as the distance between the end of an event and the start of the following event. Events in the sharp wave component whose peaks are smaller than 30 pA or separated by less than 100 ms are discarded from the analysis. To study the properties of evoked events, I inject an extra current to randomly selected 60% of B cells, at intervals of approximately 2 seconds. For each B neuron, the current is uniformly

sampled from the interval $[0, 600]$ pA and is injected for $T = 10$ ms. To avoid artifacts due to rhythmic stimulation, each stimulation time is shifted by a delay value uniformly sampled from the interval $[0, 90]$ ms. The results presented in Fig. 2.4 do not change when current injection to B cells is replaced with a depolarizing current injection to P cells, or a hyperpolarizing current injection to A cells, as far as the stimulation paradigm is strong enough to start a SWR event. Pearson correlation coefficients are computed to estimate the correlation between event amplitude and previous IEI and between event amplitude and next IEI, in both spontaneous and evoked scenarios. Only the properties of evoked events and the interval to the next (respectively previous) spontaneous SWR events are shown in the analysis of evoked events (Fig. 2.4B2, C2). The distribution of previous IEI-amplitude pairs is fitted to an exponential function $f(x) = a(1 - \exp(-bx)) + c$ with parameters a, b, and c using a non-linear least-squares method.

In the simulations with $B \to P$ depression and $P \to A$ facilitation (shown in Figs. 2.5, 2.6, 2.7), I monitor the system for 10 minutes and analyze the activity as described above. For the detection of spontaneous events in the scenario where $P \to A$ facilitation is the only plastic mechanism, the threshold to detect sharp wave peaks is adjusted to 40 pA to account for noisier events.

Effect of gabazine modulation. To mimic the effect of gabazine, which is a $GABA_A$ receptor antagonist, I reduce the inhibitory conductances g_{ij}^{IJ} ($J \in \{B, A\}$) (by the amount specified in the main text) during simulations where SWRs occur spontaneously. I analyze the system behavior during 10 minutes of simulated time before and during 10 minutes following gabazine application as follows. The LFP signal is extracted from the averaged B input to P cells. The average is taken recording the incoming B inputs from 100 randomly selected pyramidal cells. The signal is low-pass filtered at 5 Hz (Butterworth filter, order 2) and event peaks are detected as episodes exceeding 20 pA, at least 100 ms apart. Event properties are extracted as described in the previous paragraph.

2.8.5 Definition of mean f-I curves

To create the f-I curves shown in Fig. 2.11, I randomly select 50 neurons in each population, and add new neurons to the network with the same neuronal properties and incoming connectivity structure. However, I do not connect these neurons back to the network, i.e., I create copies of the selected neurons in order to study how their activity depends on the input level. To do so, I stimulate these neurons with additional constant currents of different intensities (from -100 to 200 pA, in steps of 5 pA), for $T = 20$ seconds. I distinguish periods during which the network is in either the non-SWR or the SWR state; in the latter case, depolarizing current is transiently injected to the B cells in the beginning of the simulation for the system to jump to the SWR state. All neurons in the network also receive a background current of 200 pA, as in all other simulations. I record the mean number of spikes per second, and plot this quantity against the total average input current that the neurons receive. This total current is the sum of

the external injected current, the background current, and the synaptic currents caused by incoming presynaptic activity. The gray lines in Fig. 2.11 depict single neurons' f-I curves. Additionally, the colored lines describe the mean f-I curves, averaged across neurons for each input current value. Finally, to estimate which part of the input range is more relevant to the populations in each state, I define the shaded area. The darker part represents the mean input current value (across time and neurons) seen by a neuron in a given state. The color becomes lighter until it fades at values of mean input ± one standard deviation (value computed by averaging across time and neurons).

2.8.6 Spiking network parameters

In this section, I list the parameters used in the default spiking network simulation shown in Figures 2.2, 2.3, 2.4, and 2.11.

Parameter	Value	Definition
N_P	8200	Number of pyramidal cells
N_B	135	Number PV$^+$ basket cells (B)
N_A	50	Number anti-SWR cells (A)
τ_{syn}^P	2 ms	Glutamatergic synaptic time constant
τ_{syn}^A	4 ms	GABAergic synaptic time constant (A cells)
τ_{syn}^B	1.5 ms	GABAergic synaptic time constant (B cells)
g_L	10 nS	Leak conductance
V_{rest}	-60 mV	Resting potential
V_{thr}	-50 mV	Voltage threshold
E_{rev}^P	0 mV	Excitatory reversal potential
E_{rev}^A	-70 mV	Inhibitory (A) reversal potential
E_{rev}^B	-70 mV	Inhibitory (B) reversal potential
C	200 pF	Membrane capacitance
t_{refr}^P	1 ms	Refractory period P
t_{refr}^B	1 ms	Refractory period B
t_{refr}^A	1 ms	Refractory period A
I_{BG}	200 pA	Constant background current

TABLE 2.1: Intrinsic neuronal parameters for the spiking network simulated in this chapter.

Connection	Connection probability	Conductance update (nS)	Synaptic delay (ms)
P $\rightarrow$ P	$p^{PP} = 0.01$	$g^{PP} = 0.2$	$\tau_{PP} = 1$
P $\rightarrow$ A	$p^{AP} = 0.01$	$g^{AP} = 0.2$	$\tau_{AP} = 1$
A $\rightarrow$ A	$p^{AA} = 0.6$	$g^{AA} = 4$	$\tau_{AA} = 1$
A $\rightarrow$ P	$p^{PA} = 0.6$	$g^{PA} = 6$	$\tau_{PA} = 1$
P $\rightarrow$ B	$p^{BP} = 0.2$	$g^{BP} = 0.05$	$\tau_{BP} = 1$
B $\rightarrow$ B	$p^{BB} = 0.2$	$g^{BB} = 5$	$\tau_{BB} = 1$
B $\rightarrow$ P	$p^{PB} = 0.5$	$g^{PB} = 0.7$	$\tau_{PB} = 1$
A $\rightarrow$ B	$p^{BA} = 0.6$	$g^{BA} = 7$	$\tau_{BA} = 1$
B $\rightarrow$ A	$p^{AB} = 0.2$	$g^{AB} = 8$	$\tau_{AB} = 1$

TABLE 2.2: Synaptic connectivity parameters for the spiking network simulated in this chapter. More details are provided in Sec. 2.8.2. Note that g^{AB} does not include the contribution of synaptic efficacy.

Parameter	Value	Definition
η_D	0.18	Depression rate of connections $B \rightarrow A$
τ_D	250 ms	Synaptic depression time constant of connections $B \rightarrow A$

TABLE 2.3: Synaptic depression parameters used to simulate the spiking model in this chapter. See Eq. (2.3) for details.

3 | Mean-field analysis of disinhibitory SWR networks

3.1 Introduction

In the previous chapter, I have introduced a spiking model that reproduces the main experimental features of sharp wave-ripple generation. The model exhibits sharp wave-ripple events spontaneously and in response to current injection, and the SWR dynamics is similar to experimental results. Thus, the spiking model is able, despite its simplicity, to capture the main features of the biological network of interest and to make testable experimental predictions. Additionally, it has the advantage of being defined by variables that are close to experimentally measurable quantities.

However, the large number of parameters makes the system difficult to tune by hand and impedes the understanding of the network dynamics. Why, and for which combination of parameters, does the system reproduce the experimentally observed behavior? What is the impact of one specific parameter onto the dynamics of the whole system? How robust is the network against perturbations of parameters? Answers to such questions remain elusive without a thorough mathematical analysis, that is almost impossible to perform in a spiking network like the one presented in Chapter 2.

This motivates the quest for a simpler network description, where only the average population behavior, and not the single cells' activities, is considered. Such mean-field models (also known as 'rate' models) have a long history of application in neuroscience (see e.g. [Breakspear, 2017] for a comprehensive review) and have been shown to be valuable tools for understanding the key principles governing the behavior of a biological system.

Mean-field models rely on the assumption of describing large and homogeneous populations of neurons. In this context, 'homogeneous population' means that all neurons in a population share similar intrinsic neuronal properties, receive the same amount of external input, and are coupled by statistically homogeneous connectivity - which is often assumed to be random. Taken together, these assumptions ensure that each neuron receives a large number of inputs that are, on average, the same across all neurons in a given population. Thus, by monitoring the input to a single neuron, we can obtain an accurate description of the activity at the population level. The seminal work of Wilson and Cowan shows that under these assumptions, mean-field models provide an accurate

description of the asymptotic behavior of the network [Wilson and Cowan, 1972]. The Wilson-Cowan model describes the interaction between an excitatory and an inhibitory population, whose dynamics are modeled by ordinary differential equations (ODEs). The input-output transfer functions of both populations are sigmoid functions, which allow for explicit computation of the system's stationary states. The authors show that such simple networks are well-suited to describe the stationary population behavior, but are unsuited to capture transient activity, e.g., damped oscillations during the transition to the steady state.

In this chapter, I aim at defining a mean-field model whose variables P, B, and A describe the average firing rates of the neurons in three different populations of spiking cells. To this end, a mapping is needed between the variables of the spiking and the mean-field network.

The question of how to link spiking models to mean-field descriptions has been addressed by a large number of studies. The approaches differ depending on the features of the neuronal model, current- or conductance based synapses, connectivity structure, the presence of adaptation mechanisms, the noise level in the spiking network, and the characteristics of the mean-field network (ODEs or more complex descriptions of the dynamics, and the type of transfer function used). Most derivations of mean-field models are based on LIF networks with current-based synapses (see Renart et al. [2003] for a review and Ostojic and Brunel [2011]; Montbrió et al. [2015]; Augustin et al. [2017]; Schwalger et al. [2017] for recent approaches). Only a few studies deal with the effect of voltage-dependent conductances [Ermentrout, 1994; Shriki et al., 2003; Richardson, 2004; Ostojic and Brunel, 2011; Di Volo et al., 2019]. Conductance-based networks are more complex because the voltage-dependence introduces non-linearities in the synaptic inputs (see [Gerstner et al., 2014]), but are believed to more accurately describe cortical dynamics [Borg-Graham et al., 1998; Hirsch et al., 2018; Destexhe et al., 2003].

A general drawback of most mean-field models derived from spiking models is that the transfer function cannot be expressed explicitly (as done in Wilson and Cowan [1972]), but it is derived as a function of the first- and second-order statistics of the input in the spiking network (which includes a noise component emerging from the randomness and finite size of the network). This approach is needed to quantitatively match the population rates (by including the noise contribution), but it hinders an analytical study of the mean-field model. Exceptions are the models presented by Shriki et al. [2003], where an explicit transfer function is derived under specific model assumptions (among them, that the f-I curve of the spiking network is linear), and by Fusi and Mattia [1999], where an explicit formula for the transfer function can be obtained in certain noise regimes.

The goal of this chapter is to derive a mean-field approximation of the spiking model presented in Chapter 2, which is amenable to analytical treatment. In fact, I strive to use the mean-field approach to understand how the model parameters influence the network behavior, and to unveil under which conditions one can construct networks that share the experimentally observed features presented in Chapter 2. For this reason, I focus

on networks of the Wilson-Cowan type, trying to minimize the model complexity and aiming at an explicit formulation of the transfer functions.

As a first step, I am interested in reproducing the stationary behavior of the spiking network, i.e., the coexistence of two stable states (non-SWR and SWR states) in the 3d network, in which the synaptic efficacy is clamped to an intermediate value (see Fig. 2.11). While defining the mean-field network, the bistable scenario presents the additional complication of how to define a single transfer function per population when each stationary state has a different f-I curve. Once the bistable mean-field model has been defined, we can add a synaptic depression mechanism equivalent to the one of the spiking model. This allows us to test whether the mean-field network reproduces the behavior of the spiking network upon current injection and clamping of synaptic efficacy at different holding values. The process of deriving a mean-field model from a spiking network is presented in Sec. 3.2. Furthermore, in Sec. 3.3 I derive analytical conditions for the mean-field model to display bistability, and explore the results in Sec. 3.4. Finally, I discuss drawbacks and limitations of the approach in Sec. 3.5.

Throughout the chapter, I use the population firing rates in SWR and non-SWR states as a measure to quantify whether the mean-field model successfully captures important features of the spiking network.

3.2 From spiking to mean-field model

In this section, I present an approach to define a mean-field model starting from the spiking model defined in Sec. 2.3. In Sec. 3.2.1, I define the parameters of the mean-field model. In Sec. 3.2.2, I show that the mean-field model exhibits a bistable dynamics, and reproduces the results on the effect of current injection to different populations introduced in Sec. 2.3 for the spiking model.

3.2.1 Definition of mean-field model

I define the mean-field model as a set of ordinary differential equations

$$\begin{aligned}
\tau_P \frac{\partial P}{\partial t} &= -P + f_P(W_{PP}P - W_{PB}B - W_{PA}A) \\
\tau_B \frac{\partial B}{\partial t} &= -B + f_B(W_{BP}P - W_{BB}B - W_{BA}A) \\
\tau_A \frac{\partial A}{\partial t} &= -A + f_A(W_{AP}P - W_{AB}Be - W_{AA}A) \\
\frac{\partial e}{\partial t} &= \frac{1-e}{\tau_d} - \eta Be,
\end{aligned} \tag{3.1}$$

where the first three equations describe the dynamics of the populations P, B, and A, and the fourth equation the synaptic depression mechanism, which corresponds to the synaptic depression in the spiking case. Note that e modulates the strength of the connection $B \to A$ (third equation). The transfer functions (also called activation curves) f_I, $I \in \{P, B, A\}$ describe how a population I responds to its incoming inputs. W_{IJ}

are positive values representing the average strength of the synaptic connections from population $J \in \{P, B, A\}$ to population I, and τ_d and η are the depression time constant and rate, respectively. In what follows, I briefly sketch how a mean-field network can be derived starting from the spiking network presented in Sec. 2.3. Detailed explanations are provided in Sec. 3.6.

To define the activation curves f_I I look at the stationary f-I curves of the populations in the bistable spiking network, i.e. with synaptic efficacy $e_{IJ} = 0.5$ (see Fig. 2.11). For asynchronously firing neurons, it has been shown that single neurons' f-I curves are sufficient to define the populations' activation curves [Brunel, 2000b; Brunel and Wang, 2003; Gerstner et al., 2014]. A complication in the current setup is that, as two stationary states exist, one activation function per population needs to be defined starting from two sets of f-I curves (one per state). To this end, I observe that the input level to a given population is different in the non-SWR and SWR states (see Fig. 2.11). To define the activation functions, a softplus function [Dugas et al., 2001; Glorot et al., 2011], defined as $f_I(x) = \ln\{1+\exp[k_I(x+s_I)]\}$, is fitted to the sum of the two f-I curves in both stable states, weighted depending on where most inputs arrive in either state (see Sec. 3.6.1). This allows us to compute k_I and s_I for $I \in \{P, B, A\}$. To define the three activation functions $f_P(x)$, $f_B(x)$, and $f_A(x)$, I additionally include in the mean-field input the $I = 200$ pA background current that all neurons in the spiking network receive. In other words, I define $t_I = s_I + 200$ as the threshold of $f_I(x)$; no extra background current is injected to the populations in the mean-field model. Thus, the softplus functions used in the mean-field simulations are

$$f_I(x) = \ln\{1 + \exp[k_I(x + t_I)]\}, \quad I \in \{P, B, A\}. \tag{3.2}$$

In addition, the mean connection strengths W_{IJ} need to be defined. Intuitively, they should depend on the mean number of synapses to a given postsynaptic cell, and on the average strength of a single postsynaptic potential (PSP). In turn, a PSP varies depending on the unitary conductance update and the driving force, i.e., the difference between the reversal potential for a given synapse type (excitatory or inhibitory) and the mean membrane potential value of the postsynaptic population. The existence of two stationary states in the spiking network complicates the estimation of the mean membrane potential values because these values differ across different states (see Fig. 3.12). For this reason, I consider these values as free parameters of the system and run an optimization procedure to find the values yielding the best match between population firing rates in the spiking and in the mean-field model. More details on the optimization procedure are provided in Sec. 3.6.1.

The parameters τ_P, τ_B, and τ_A in Eq. (3.1) set the time constants of the population dynamics. It is well known that no correspondence can be drawn between the membrane time constants of the spiking network and the population time constants [Abbott, 1994; Dayan and Abbott, 2001; Gerstner et al., 2014]. As a result, using the mean-field model as an approximation of the spiking model can at most hold in the stationary, but not in

the transient, case. This is a known and important limitation of mean-field models, even though recent approaches that go beyond a Wilson-Cowan formulation tried to address this problem [Montbrió et al., 2015; Schwalger et al., 2017]. I set the population time constants in Eq. (3.1) to $\tau_P = 3$ ms, $\tau_B = 2$ ms, and $\tau_A = 6$ ms. These values are biologically plausible [Wilson and Cowan, 1972; Chenkov et al., 2017] and account for the fact that B cells are assumed to be fast interneurons; I additionally assume that A cells are slower interneurons. Note, however, that the asymptotic dynamics is largely independent on the choice of the time constants.

The last ingredient needed to create the mean-field model envisioned in Eq. (3.1) is the definition of the synaptic depression equation. It can be directly derived from the spiking case (Eq. (2.3)) by averaging over realizations (i.e. $e = \overline{e_{IJ}}$, where e_{IJ} is the average of the synaptic efficacies e_{ij} of synapses $j \to i$), under the assumption of considering a large number of presynaptic spikes. In this scenario, the synaptic efficacy evolves as described in Eq. (3.1), with $\eta = \eta_D$ and $\tau_d = \tau_D$.

3.2.2 Comparison between mean-field and spiking model

Now that all the elements of Eq. (3.1) have been defined, we can test the behavior of the mean-field model and compare it to the simulations of the spiking model. First, I clamp the synaptic efficacy at different values. Figure 3.1 shows that for $e = 0.5$ two stable states coexist (compare to Fig. 2.11 for the spiking model). Current injection to the P population is sufficient to switch from the non-SWR state to the SWR state. Thanks to the choice of the mean membrane potential values (see Sec. 3.6.1), the quantitative match between firing-rate values in spiking and mean-field models is accurate, with deviations $< 1\ \mathrm{s}^{-1}$. Additionally, setting $e = 0.2$ terminates the SWR state because in this scenario the A population receives too little inhibition from the B population and can thus restore its firing rate to non-SWR levels.

Interestingly, even if the synaptic efficacy is set to a large value, e.g., $e = 0.8$, the system does not jump to the SWR state. This behavior is a consequence of the lack of noise in the network. In contrast, in the spiking network, fluctuations can trigger a jump to the SWR state (see Fig. 2.11). If the system is fully deterministic, no jumps are expected as far as the change in synaptic efficacy preserves the network bistability. This implies that the mean-field network is not able to reproduce fluctuation-driven spontaneous SWRs observed in the spiking network and in experiments.

Nevertheless, we can compare the time-dependent behavior upon current injection in both models, when the synaptic efficacy dynamics is not clamped, but can evolve as stated in Eq. (3.1). Figure 3.2 shows that the mean-field model behavior is comparable to the one of the spiking model, with P and B transiently increasing and A transiently decreasing upon current injection (in all scenarios, in which current is injected to a different population). The extreme values of the firing rates are those observed in the bistable case and are thus close to the spiking values. As in the spiking case, the depression mechanism brings the system from a SWR state back to the non-SWR state. Thus, we

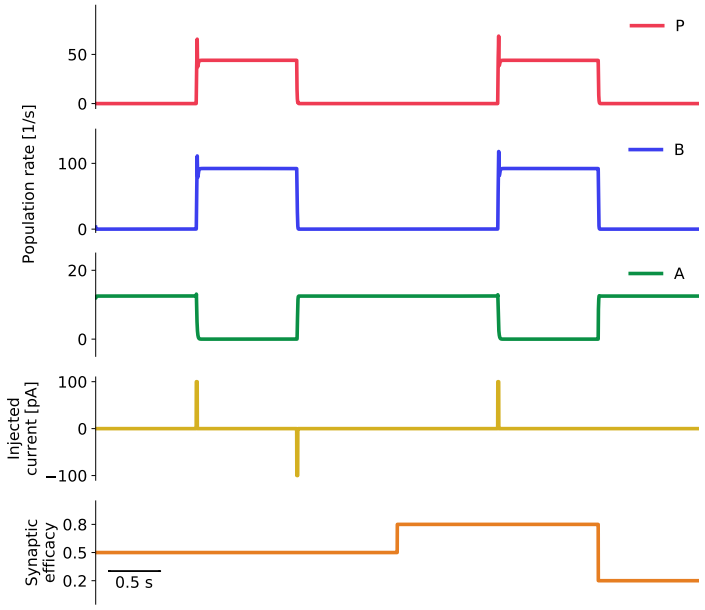

FIGURE 3.1: **Mean-field network with fitted softplus f-I curves is bistable for clamped depression values.** The plot shows, from top to bottom: population rates of P, B, and A cells, injected current, and value of synaptic efficacy. When the synaptic efficacy is clamped at $e = 0.5$, two stable states are present in the network (left). Positive current injection to the P population triggers the switch to the SWR state, while negative current terminates it. These results are comparable to what has been shown in Fig. 2.11 for the spiking model. The population firing rate values are matched in both networks due to the optimization of the mean membrane voltages V_I (see text for details). Population rates are, in the non-SWR state: $P = 0$ s^{-1}, $B = 0$ s^{-1}, $A = 12.5$ s^{-1}, and in the SWR state: $P = 44.0$ s^{-1}, $B = 92.2$ s^{-1}, $A = 0$ s^{-1}, see Fig. 2.11 for comparison with spiking values. Differently from the spiking model, the mean-field model does not jump to the SWR state for $e = 0.8$ because of its noise-free nature. When a positive current is injected, so that the system jumps to the SWR state, the event can be terminated by lowering the synaptic efficacy to $e = 0.2$. Network parameters are: $W_{PP} = 1.72$, $W_{BP} = 8.86$, $W_{AP} = 1.72$, $W_{PB} = 1.24$, $W_{BB} = 3.24$, $W_{AB} = 5.66$, $W_{PA} = 12.60$, $W_{BA} = 13.44$, $W_{AA} = 8.40$, $k_P = 0.47$, $k_B = 0.41$, $k_A = 0.48$, $t_P = 131.66$, $t_B = 131.96$, $t_A = 131.09$, $\tau_P = 3$ ms, $\tau_B = 2$ ms, $\tau_A = 6$ ms, $\eta = 0.18$, $\tau_d = 250$ ms (these two parameters do not play a role as the synaptic efficacy is clamped at $e = 0.5$).

can conclude that the mean-field model accurately reproduces the stationary behavior of the spiking network.

The mean-field model seems thus a suitable tool to describe the population dynamics of the spiking model. Yet this approach faces some unavoidable limitations. First, as mentioned above, the model is unsuited for describing the transient dynamics of the spiking network. A direct consequence is the lack of ripple oscillations in the mean-field simulation (compare Fig. 3.2 with Fig. 2.2). This also implies that even matching the input currents used in Figs. 2.2, 2.11, and 3.1, 3.2 in spiking and mean-field models, respectively, would not lead to the same behavior in the initial phase of a SWR event. Indeed, Figures 3.1 and 3.2 show that current injection (which would correspond to the stimulation of all neurons in the spiking scenario) generates an overshoot. In the spiking model, the first oscillation in the population rate could be interpreted as an overshoot

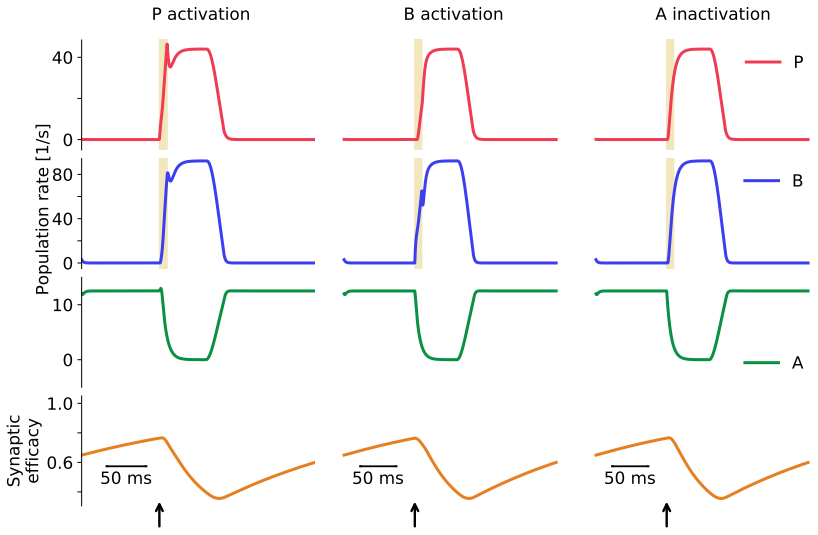

FIGURE 3.2: **Mean-field network simulations with fitted softplus f-I curves displays SWR-like events.** From top to bottom: P, B, and A firing rates, and level of synaptic efficacy. In each column, current is injected to a different population (P, B, A from left to right). In all cases, switching on the current (yellow regions) makes the system switch to the SWR state. The depression mechanism (bottom) brings the system back to rest. Results are comparable with the population firing rates of the spiking network presented in Fig. 2.2. Network parameters are listed in Fig. 3.1.

(see Fig. 2.2), but its size can be better controlled by the stimulation paradigm (i.e., number of stimulated cells and current value to each cell).

Second, some of the mean-field assumptions are violated. In fact, the number of cells in each population is not really large (as few as 50 cells belong to the A population), and the SWR state is not asynchronous, as the population activity is characterized by ripple-like events. In this respect, it is quite surprising to see that the mean-field model can achieve such a good match with the spiking model.

Third, the process of approximating the spiking network with a mean-field model is not univocal, as it depends on the choice of τ_P, τ_B, and τ_A (population time constants) and V_P, V_B, and V_A (mean membrane potential values used to define the connection strengths W_{IJ}). Overall, we can conclude that the two models exhibit similar features, but should not be considered equivalent.

The crucial advantage of the mean-field model over its spiking formulation is that it can be used to analytically predict, as a function of the mean-field parameters, when the network exhibits bistability. In this way, we can understand the influence of each parameter on the behavior of the system and extend the range of bistable solutions to parameters yet untested in the spiking network. The analysis is presented in the next section.

3.3 Mean-field model analysis

I now focus on studying under which conditions the system in Eq. (3.1) with frozen synaptic efficacy ($e = const$) is bistable. I have previously shown (in Sec. 3.2) that this happens for intermediate values of synaptic efficacy, and I choose $e = 0.5$ as default value for e in what follows. The fixed synaptic efficacy value can be thought of as a factor scaling the W_{AB} connection value. For simplicity, I consider in what follows the resulting 3-dimensional system with halved W_{AB} and $e = 1$, i.e., I study the system

$$\tau_P \frac{\partial P}{\partial t} = -P + f_P(W_{PP}P - W_{PB}B - W_{PA}A) \tag{3.3a}$$

$$\tau_B \frac{\partial B}{\partial t} = -B + f_B(W_{BP}P - W_{BB}B - W_{BA}A) \tag{3.3b}$$

$$\tau_A \frac{\partial A}{\partial t} = -A + f_A(W_{AP}P - W_{AB}B - W_{AA}A) \tag{3.3c}$$

where the e contribution is absorbed into W_{AB}. To study conditions for bistability, it is useful to approximate the softplus functions defined in Eq. (3.2) using the threshold linear approximations

$$f_P(x) = \begin{cases} k_P(x + t_P) & \text{if } x \geq -t_P \\ 0 & \text{else} \end{cases} \tag{3.4}$$

$$f_B(x) = \begin{cases} k_B(x + t_B) & \text{if } x \geq -t_B \\ 0 & \text{else} \end{cases} \tag{3.5}$$

$$f_A(x) = \begin{cases} k_A(x + t_A) & \text{if } x \geq -t_A \\ 0 & \text{else} \end{cases} \tag{3.6}$$

The terms t_I can be interpreted as activity thresholds, and the terms k_I define the slopes of the activation functions. I set $k_I > 0$ and $t_I > 0$, $\forall I \in \{P, B, A\}$. Equations (3.4), (3.5), and (3.6) are good approximations of Eq. (3.2) when the exponential term becomes dominant in the argument of the logarithm, i.e., when $\exp(k_I(x+t_I)) \gg 1$. The threshold linear activation functions allow us to explicitly compute the nullclines and fixed points of the system presented in Eq. (3.3), as explained below.

The analysis I present here aims at studying the steady-state behavior of the mean-field system (Eq. (3.1)) using standard dynamical systems theory. The overarching aim is to understand for which parameter choices the system displays bistability. Additionally, to accurately reproduce experiments (see Sec. 1.2.3) and results obtained in simulations of the spiking network (see Sec. 2.3), positive input currents to P and B, and negative input currents to A should be sufficient for the system to switch to the SWR state. As discussed in Sec. 2.3.1, this problem is non-trivial because, depending on the relative strength of the pathways, the activation patterns upon population stimulation might be

different. Thus, we are looking for conditions for the network to be *both* bistable and with the right behavior upon population stimulation. Hence, I study, as a function of the mean-field parameters, the existence and stability of three fixed points (FPs), two stable and one unstable. One stable FP should correspond to the non-SWR state, with A population active, P population only mildly active, and B population almost inactive (FP0 in what follows). The other stable FP is the SWR state, with P and B strongly active and A almost inactive (FP1 in what follows). The unstable FP corresponds to an intermediate FP between FP1 and FP0, which can be interpreted as a threshold for the system to switch between the two states (and is called FP2). All FPs define firing rates and should therefore be non-negative. To simplify the analysis, I will assume that $B = 0$ in FP0 and $A = 0$ in FP1. Additionally, I aim at enforcing the conditions on the strength of the pathways introduced in Sec. 2.3.1.

I sketch below the main steps of the analysis and refer to Sec. 3.6.3 for details. The analysis is made out of three main components: the first part deals with the derivation of the system nullclines, the second part describes how geometrical considerations can be used to define conditions for the nullclines to intersect and yield 3 FPs, and the final part deals with the FPs definition and stability analysis. The outcome of the analysis is a set of relatively simple conditions, which can be simultaneously enforced to find a region in parameter space where the network is bistable and the requirements presented in Sec. 2.3.1 are fulfilled. This method allows us to understand the influence of different parameters on the network behavior. Additionally, it provides insights on the important ingredients needed to create networks displaying the features described in Sec. 2.3.

3.3.1 Nullclines and fixed-points computation

To simplify the analysis, I assume that $B = 0$ in the non-SWR state (FP0) and $A = 0$ in the SWR state (FP1), i.e., I assume that B and A receive subthreshold inputs in these states (see Eqs. (3.5), (3.6)). This assumption is justified by the fact that, both in experimental results and spiking model (see Sec. 1.2.3 and Sec. 2.3), the cell populations are almost inactive (i.e. cells fire only a few spikes per second). This choice simplifies the study of the nullclines because it allows us to perform the analysis in two 2-dimensional subspaces, instead of studying the nullclines in the full 3-dimensional space. In fact, I first restrict the analysis to the 2d space called 'B-P', assuming that $A > 0$ (i.e. I exclude the case in which $A = 0$), and study the intersections of the P and B nullclines yielding the fixed points FP0 and FP2. Afterward, I replicate the analysis in the A-P space, where I assume that $B > 0$ and compute and visualize the fixed points FP1 and FP2.

3.3.1.1 Nullclines in the B-P space

I first present the main results of the analysis in the B-P space (complete derivations are presented in Sec. 3.6.3.1). To compute the nullclines I consider the system presented in Eq. (3.3) when the derivatives on the left-hand side are set to zero. First, I set $\frac{\partial A}{\partial t} = 0$

in Eq. (3.3c). Because of the assumption $A > 0$, the equation can be solved for A and used to compute the P and B nullclines in this space. As the activation functions of the system are threshold linear functions, assuming $A > 0$ imposes that the input of Eq. (3.6) is larger than the threshold. This defines the condition

$$W_{AP}P - W_{AB}B - W_{AA}A \geq -t_A \Leftrightarrow B < \frac{W_{AP}P + t_A}{W_{AB}}, \tag{3.7}$$

which describes a region where nullclines and FPs are well-defined in the B-P space. Similarly, solving Eq. (3.3b) for B, to define the non-zero portion of the B nullcline, is only possible for input values of Eq. (3.5) above the threshold. Similar considerations hold for the P nullcline.

The non-zero part of the B nullcline is defined as

$$\begin{aligned}\underbrace{\left(1 + k_B W_{BB} - \frac{k_B k_A W_{BA} W_{AB}}{1 + k_A W_{AA}}\right)}_{=:l} B = &\underbrace{\left(k_B W_{BP} - \frac{k_B k_A W_{BA} W_{AP}}{1 + k_A W_{AA}}\right)}_{=:\gamma} P \\ &+ \underbrace{\left(k_B \left(-\frac{k_A t_A W_{BA}}{1 + k_A W_{AA}} + t_B\right)\right)}_{=:m},\end{aligned} \tag{3.8}$$

so that the B nullcline can be defined, assuming $l \neq 0$, as

$$B = \begin{cases} \frac{1}{l}(\gamma P + m) & \text{if } W_{BP}P - W_{BB}B - W_{BA}A \geq -t_B \\ 0 & \text{else} \end{cases} \tag{3.9}$$

An interesting feature of this analysis is that the terms γ, l, and m can be linked to strengths of connections in the mean-field network. In fact, γ can be interpreted as the difference in strength between the pathway $P \to B$ and the pathway $P \to \overset{\curvearrowleft}{A} \to B$ (which includes the recurrent connection W_{AA}). Therefore, we can incorporate the requirements on pathway strengths discussed in Sec. 2.3.1 into the bistability conditions by imposing constraints on the sign of the newly defined terms γ, l, and m. This constrains the set of bistable networks we focus on, while allowing us to derive conditions that yield a network configuration where both bistability and behavior upon current injection are enforced. For example, to comply with requirement 2 in Sec. 2.3.1, we need $\gamma > 0$. The term l can be thought of as the difference in strength between the recurrent connectivity in B and the pathway $B \to \overset{\curvearrowleft}{A} \to B$. The term m can be interpreted as the offset between a baseline input arriving at B through the pathway $A \to B$ and the threshold in B. The terms l and m have no sign constraints.

Analogously, the non-zero portion of the P nullcline can be written as

$$\underbrace{\left(\frac{k_P k_A W_{PA} W_{AB}}{1 + k_A W_{AA}} - k_P W_{PB}\right)}_{=:c} B = \underbrace{\left(1 - k_P W_{PP} + \frac{k_P k_A W_{PA} W_{AP}}{1 + k_A W_{AA}}\right)}_{=:g} P + \underbrace{\left(k_P \left(\frac{k_A t_A W_{PA}}{1 + k_A W_{AA}} - t_P\right)\right)}_{=:h}. \tag{3.10}$$

The P nullcline is then defined as

$$P = \begin{cases} \frac{1}{g}(cB - h) & \text{if } W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P \\ 0 & \text{else} \end{cases} \tag{3.11}$$

The term c can be interpreted as the difference in strength between the pathway $B \to \overset{\curvearrowleft}{A} \to P$ and the pathway $B \to P$. We enforce requirement 3 in Sec. 2.3.1 by setting $c > 0$. The term g can be thought of as the summed strength of the recurrent connectivity in P and the pathway $P \to \overset{\curvearrowleft}{A} \to P$. We assume that $1 - k_P W_{PP} > 0$, i.e., we assume that recurrent excitatory connections are small enough to avoid activity explosion, which implies $g > 0$. As a result, the non-zero portion of the P nullcline defined in Eq. (3.11) has a positive slope. The term h can be interpreted as the offset between a baseline input arriving at P through the pathway $A \to P$ and the threshold in P, and its sign is unconstrained.

In Figure 3.3A, we visualize the B and P nullclines in the B-P space, for the mean-field model presented in Fig. 3.1. Both nullclines are only defined inside the green shaded portion, defined by Eq. (3.7). The B nullcline (blue line) is non-zero inside the half-plane hatched with blue lines, and is defined as $B = 0$ outside of it (to satisfy Eq. (3.5)). Similarly, the P nullcline (red line) is non-zero inside the half-plane hatched with red circles and equals $P = 0$ outside of it (to satisfy Eq. (3.4)).

Now that the P and B nullclines have been defined, I impose conditions for them to intersect in two points in the first quadrant (because P and B represent firing rates, they should be non-negative). The non-SWR fixed point (FP0) should have $B = 0$ by construction, i.e. it should belong to the zero branch of the B nullcline; on the other hand, FP2 should belong to the non-zero branch of the B nullcline.

The conditions yielding 2 intersections of the nullclines include a condition on their slopes and one on their intersections with the line $B = 0$ being in the correct order. I refer to Sec. 3.6.3.1 for a complete list of conditions.

3.3.1.2 Nullclines in the A-P space

The analysis I presented in Sec. 3.3.1.1 can be replicated in the A-P space. The main difference is that the conditions on the strengths of the network pathways impose a different structure to the nullclines. As in the previous section, I report here the main results and refer to Sec. 3.6.3.2 for complete derivations.

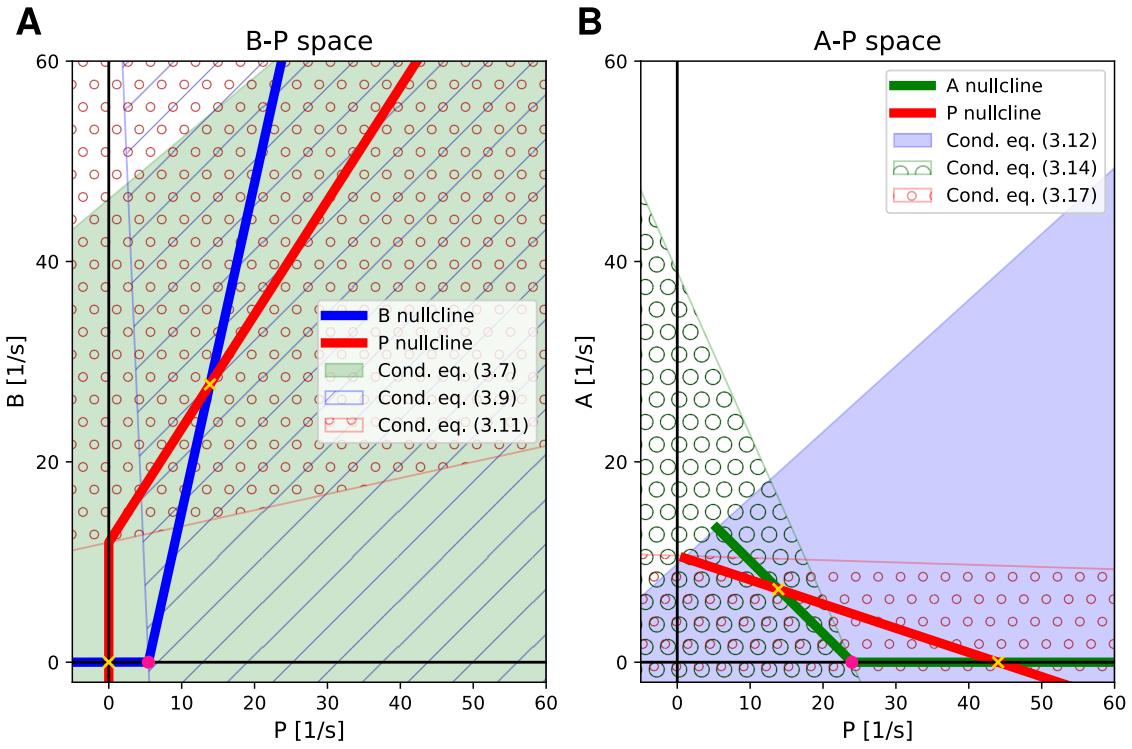

FIGURE 3.3: **Nullclines of the mean-field model derived from the spiking network.** The plots show the nullclines in the 2d subspaces B-P (**A**) and A-P (**B**), for the network simulated in Fig. 3.1 (with $e = 0.5$). Network parameters are: $W_{PP} = 1.72$, $W_{BP} = 8.86$, $W_{AP} = 1.72$, $W_{PB} = 1.24$, $W_{BB} = 3.24$, $W_{BA} = 13.44$, $W_{AB} = 2.83$ (note that this value has been halved to include the contribution of the synaptic efficacy variable $e = 0.5$), $W_{PA} = 12.60$, $W_{AA} = 8.40$, $k_P = 0.47$, $k_B = 0.41$, $k_A = 0.48$, $t_P = 131.66$, $t_B = 131.96$, $t_A = 131.09$. **A)** P (red) and B (blue) nullclines in the B-P space. Green shaded area represents the portion where Eq. (3.7) holds; nullclines are only defined inside this area. The area hatched with blue lines defines the portion where the B nullcline is non-zero (first branch of Eq. (3.9)), and the area hatched with red circles represents the portion where the P nullcline is non-zero (first branch of Eq. (3.11)). Outside of these regions, nullclines are zero. Gold crosses are the FPs of the system ($FP0$ with $P_0 = 0$, $B_0 = 0$; and $FP2$ with $P_2 > 0$, $B_2 > 0$). The pink dot represents the point $-m/\gamma$. **B)** P (red) and A (green) nullclines in the A-P space. Blue shaded area represents the portion where Eq. (3.12) holds; nullclines are only defined inside this area. The area hatched with green circles defines the portion where the A nullcline is non-zero (first branch of Eq. (3.14)), and the area hatched with red circles represents the portion where the P nullcline is non-zero (first branch of Eq. (3.17)). Outside of these regions, nullclines are zero. Gold crosses are the FPs of the system ($FP1$ with $P_1 = -h'/g'$, $A_1 = 0$; and $FP2$ with $P_2 > 0$, $B_2 > 0$), and the pink dot represents the point $-m'/\gamma'$.

I restrict the analysis to the 2d space A-P, assuming that $B > 0$ (i.e. I exclude the case in which $B = 0$). I set $\frac{\partial B}{\partial t} = 0$ in Eq. (3.3b). Because of the $B > 0$ assumption, the equation can be solved for B and used to compute the P and A nullclines in this space. The assumption $B > 0$ imposes that the input of Eq. (3.5) is larger than the threshold. This yields the condition

$$W_{BP}P - W_{BB}B - W_{BA}A \geq -t_B \Leftrightarrow A < \frac{W_{BP}P + t_B}{W_{BA}}. \tag{3.12}$$

This condition describes the region where nullclines and FPs are well-defined in the A-P space.

The non-zero portion of the A nullcline can be written as

$$\begin{aligned}\underbrace{\left(1 + k_A W_{AA} - \frac{k_A k_B W_{AB} W_{BA}}{1 + k_B W_{BB}}\right)}_{=:l'} A = &\underbrace{\left(k_A W_{AP} - \frac{k_A k_B W_{AB} W_{BP}}{1 + k_B W_{BB}}\right)}_{=:\gamma'} P \\ &+ \underbrace{\left(k_A \left(-\frac{k_B t_B W_{AB}}{1 + k_B W_{BB}} + t_A\right)\right)}_{=:m'},\end{aligned} \tag{3.13}$$

so that the A nullcline can be defined, assuming $l' \neq 0$, as

$$A = \begin{cases} \frac{1}{l'}(\gamma' P + m') & \text{if } W_{AP}P - W_{AB}B - W_{AA}A \geq -t_A \\ 0 & \text{else} \end{cases} \tag{3.14}$$

The term γ' can be interpreted as the difference in strength between the pathway $P \to A$ and the pathway $P \to \overset{\curvearrowleft}{B} \to A$. To comply with requirement 1 in Sec. 2.3.1, we need $\gamma' < 0$. The term l' can be thought of as the difference in strength between the recurrent connectivity in A and the pathway $A \to \overset{\curvearrowleft}{B} \to A$. Note that

$$l' = \frac{l(1 + k_A W_{AA})}{1 + k_B W_{BB}}, \tag{3.15}$$

and l' will therefore have the same sign as l. The term m' can be interpreted as the offset between a baseline input arriving to A through the pathway $B \to A$ and the threshold in A. The terms l' and m' can have a positive or a negative sign.

The non-zero branch of the P nullcline can be defined as

$$\begin{aligned}\underbrace{\left(\frac{k_P k_B W_{PB} W_{BA}}{1 + k_B W_{BB}} - k_P W_{PA}\right)}_{=:c'} A = &\underbrace{\left(1 - k_P W_{PP} + \frac{k_P k_B W_{PB} W_{BP}}{1 + k_B W_{BB}}\right)}_{=:g'} P \\ &+ \underbrace{\left(k_P \left(\frac{k_B t_B W_{PB}}{1 + k_B W_{BB}} - t_P\right)\right)}_{=:h'},\end{aligned} \tag{3.16}$$

so that the P nullcline can be defined as

$$P = \begin{cases} \frac{1}{g'}(c'A - h') & \text{if } W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P \\ 0 & \text{else} \end{cases} \tag{3.17}$$

The term c' can be interpreted as the difference in strength between the pathway $A \to \overset{\curvearrowleft}{B} \to P$ and the pathway $A \to P$. To comply with requirement 5 in Sec. 2.3.1, we need $c' < 0$. The term g' can be thought of as the summed strength of the recurrent connectivity in P and the pathway $P \to \overset{\curvearrowleft}{B} \to P$. As we have assumed in Sec. 3.3.1.1 that $1 - k_P W_{PP} > 0$, it follows that $g' > 0$. The term h' can be interpreted as the offset between a baseline input arriving to P through the pathway $B \to P$ and the threshold in P. At this point, we do not impose any sign constraint on h'.

Figure 3.3B shows the P and A nullclines in the A-P space for the mean-field model presented in Eq. (3.3) (and in Fig. 3.1 when $e = 0.5$). Both nullclines are only defined inside the blue portion, defined by Eq. (3.12). The A nullcline (green line) is non-zero inside the half-plane hatched with green circles and zero outside of it. The P nullcline is non-zero inside the half-plane hatched with red circles, and would be $P = 0$ outside of it. However, the zero branch of the P nullcline is not defined because it lies outside of the domain of definition of the nullclines (i.e. outside of the shaded blue region).

I now briefly turn to the conditions for the nullclines to intersect in two points. By definition, the SWR fixed point (FP1) should have $A = 0$, i.e., it should belong to the zero branch of the A nullcline. Its P coordinate can be zero, or positive. For FP2, both coordinates should be positive. Thus, the nullclines should have two intersections in the first quadrant; this can be enforced by a condition on the slopes, and by a condition on the intersections of the nullclines with the line $A = 0$ being properly ordered. The full list of conditions is presented in Sec. 3.6.3.2.

3.3.1.3 Fixed points' equations and stability

The last part of the analysis deals with the stability of the FPs: for the system to be bistable, the non-SWR and SWR fixed points (FP0 and FP1) should be stable, and FP2 (representing the threshold for switching) should be unstable. First, the FP equations are easily derived as intersections of nullclines (see Sec. 3.6.3.3 for details):

Non-SWR state (FP0):

$$\begin{aligned} P_0 &= \begin{cases} -\dfrac{h}{g} & \text{if } h < 0 \\ 0 & \text{else} \end{cases} \\ B_0 &= 0 \\ A_0 &= \frac{k_A W_{AP} P_0 + k_A t_A}{1 + k_A W_{AA}} \end{aligned} \tag{3.18}$$

Note that g (defined in Eq. (3.10)) is positive, and that B_0 is zero by construction. Because $P_0 \geq 0$ and $k_A > 0$, $t_A > 0$, it follows that A_0 is always positive.

SWR state (FP1):

$$\begin{aligned} P_1 &= -\frac{h'}{g'} \\ B_1 &= \frac{k_B W_{BP} P_1 + k_B t_B}{1 + k_B W_{BB}} \\ A_1 &= 0 \end{aligned} \tag{3.19}$$

A_1 is zero by construction. Because $P_1 \geq 0$ and $k_B > 0$, $t_B > 0$, it follows that B_1 is always positive.

Middle FP (FP2):

$$\begin{aligned} P_2 &= \frac{cm - lh}{lg - c\gamma} = \frac{c'm' - l'h'}{l'g' - c'\gamma'} \\ B_2 &= \frac{gm - \gamma h}{gl - \gamma c} \\ A_2 &= \frac{g'm' - \gamma' h'}{g'l' - \gamma' c'} \end{aligned} \tag{3.20}$$

By looking at the equations defining the variables g, c, l, etc., we note that only the middle FP (FP2) depends on W_{AB}, as shown in Figure 3.4. I remind the reader that W_{AB} represents the strength of the connection $B \to A$, which is the one modulated by the synaptic efficacy variable (see Eq. (3.1)). Thus, the fact that the non-SWR and SWR FPs are independent of W_{AB} implies that the level of synaptic efficacy chosen to study the mean-field bistable network does not affect the FP values (see Fig. 3.1). This is an interesting property of the system because it shows that the effect of clamping the synaptic efficacy e at different levels moves the threshold (the middle fixed point FP2), i.e., it modifies how easy it is to switch between states in either direction, without altering the states themselves. Note that the picture is different in the spiking network, in which the level of depression also influences the level of noise in the system (see Fig. 2.11).

To define conditions for the FPs to have the right stability, I linearize the system around each fixed point and compute the stability looking at the sign of the eigenvalues of the Jacobian matrix. This is a standard method to study the stability of FPs in dynamical systems, and the I refer to Strogatz [2018] for a detailed introduction to the method, and to Sec. 3.6.3.3 for the derivation in our network of interest. From the assumptions (in Sec. 3.3.1.1 and Sec. 3.3.1.2) $g > 0$ and $g' > 0$, it follows that both FP0 and FP1 are stable. The stability of the middle FP is more complex because the eigenvalues of the Jacobian matrix associated with this FP (termed J_{middle}) can be calculated explicitly, but are not insightful due to their complicated form. For this reason, I resort to numerical methods to impose the instability of this point (i.e. there exist at least one positive and one negative eigenvalue). Numerical computation shows the existence of one positive and two negative eigenvalues for the network presented in Fig. 3.3.

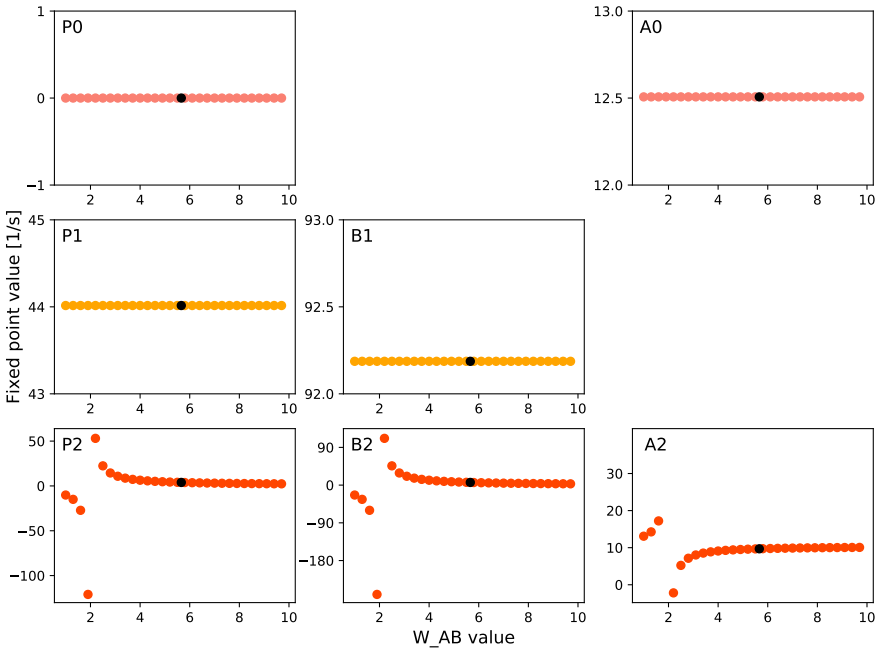

FIGURE 3.4: **Fixed points' dependence on $\mathbf{W_{AB}}$.** Fixed point values are plotted as a function of W_{AB}. B_0 and A_1 are not displayed as they are set to zero by construction. Note that only the coordinates of the middle FP depend on W_{AB} (also note the divergence at $W_{AB} \approx 2$). The independence of FP0 and FP1 on W_{AB} holds for any parameter choice; in this example, the parameters are the same as in Fig. 3.1 (the black dots represent the value of W_{AB}). The nullclines for this network are displayed in Fig. 3.3 and the simulations in Figs. 3.1 and 3.2.

3.3.1.4 Summary of conditions

I summarize below the conditions to be enforced for the system to have three FPs with the right stability (FP0 and FP1 stable, FP2 unstable). Conditions 1 (on c, γ) and 2-5 are derived in the B-P space (see Sec. 3.6.3.1) conditions 1 (on c', γ'), and 6-8 are derived in the A-P space (see Sec. 3.6.3.2), condition 9 has been imposed to fulfill experimental results on the P population being more active in SWR states than in non-SWR states (see Sec. 1.2.3), and condition 10 is derived in Sec. 3.6.3.3.

1. $c > 0$, $\gamma > 0$, $c' < 0$, $\gamma' < 0$ (requirements on pathways strength)
2. $1 - k_P W_{PP} > 0$ (small excitatory recurrent connections; it implies $g > 0$, $g' > 0$)
3. if $l > 0$, impose $gl - \gamma c < 0$ (slope condition in B-P space, always holds for $l \leq 0$)
4. if $l \geq 0$, it should hold $0 < -\frac{m}{\gamma}$. If $l < 0$ and $h < 0$, we need to impose $-\frac{h}{g} < -\frac{m}{\gamma}$, whereas if $l < 0$ and $h \geq 0$, it should hold that $0 < -\frac{m}{\gamma}$ and that $cm - lh < 0$ (ordered intersections at $B = 0$, non-negative FPs)
5. $B_2 < \frac{W_{AP} P_2 + t_A}{W_{AB}}$ (for Eq. (3.7) to be fulfilled at FP2)
6. if $l' > 0$, impose $g'l' - \gamma' c' < 0$ (slope condition in A-P space, always holds if $l' \leq 0$)
7. $-\frac{m'}{\gamma'} < -\frac{h'}{g'}$ and $c'm' - l'h' < 0$ (for ordered intersections at $A = 0$ and non-negative FPs; it implies $h' < 0$). If $l' \geq 0$, it additionally holds that $m' > 0$
8. $A_2 < \frac{W_{BP} P_2 + t_B}{W_{BA}}$ (for Eq. (3.12) being fulfilled at FP2)
9. $hg' - h'g > 0$ (to have $P_0 < P_1$)

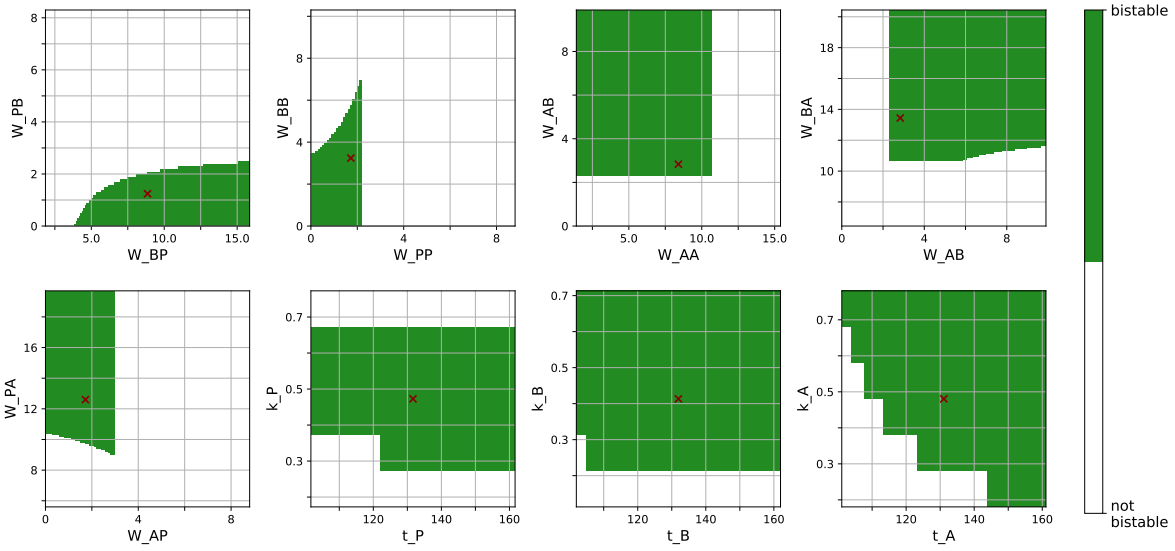

FIGURE 3.5: **Multiple network configurations support bistability.** For each plot, the green area describes where conditions 1 - 10 are fulfilled, and thus the network is bistable with the right activation patterns upon population stimulation. Each subplot is obtained by enforcing conditions 1 - 10 as two parameters are varied (x- and y-axis, respectively). The red crosses depict the values of the parameters in the network derived from the spiking model defined in Sec. 2.3 (see Fig. 3.3 for a complete list of parameters).

10. J_{middle} has at least one positive eigenvalue (numerical)

By simultaneously enforcing these conditions, we can restrict the parameter space to regions in which the system is bistable and the conditions on pathway strength guarantee that, upon population activation, the system responds as the network presented in Fig. 3.2 (i.e., as shown by experimental results, see Sec. 1.2.3).

3.4 Exploring the space of mean-field models

In Section 3.3, I have derived quasi-analytical conditions to create 3-dimensional mean-field models that are bistable and show the right activation pattern upon current injection. This methodology will allow us to understand the principles needed to construct mean-field models that exhibit SWR-like behavior comparable to experiments. In Sec. 3.4.1, I explore the parameter space where all conditions (see Sec. 3.3.1.4) are fulfilled and show that mean-field models inside this region display a behavior similar to the network in Fig. 3.1. To better understand the importance of enforcing the conditions on the pathway strength (see Sec. 2.3.1), I show in Sec. 3.4.2 an example network where these conditions are partially violated. Finally, in Sec. 3.4.3, I show that new spiking models can be created, that have a similar behavior to the network presented in Chapter 2. Even though this procedure is only possible to a certain extent, the steps presented in this chapter significantly advance our theoretical understanding of the principles governing 3-dimensional networks with SWR-like behavior.

3.4.1 Inside the region where conditions are fulfilled

The conditions presented in Sec. 3.3.1.4 are the building blocks that define the networks we are interested in. It is difficult to imagine how the space where all conditions are fulfilled looks like because of the large number of parameters (15: 9 connection strengths W_{IJ}, 3 thresholds t_I and 3 slopes k_I, see Eq. (3.1)) and conditions (10, see Sec. 3.3.1.4). However, Fig. 3.5 shows that for many parameters combinations around the starting point, i.e., the mean-field model that approximates the spiking model defined in Chapter 2 (see Sec. 3.2 for its derivation), all conditions are simultaneously fulfilled (the green regions are the projections of this space in 2d subspaces).

The analysis I developed in Sec. 3.3 is valid for the stationary state of the system. However, it remains to be verified whether, upon transient current injection and depression mechanism, the networks show the desired behavior. To this end, I randomly select one point in the region and check that: i) the nullclines of the system intersect in 3 fixed points; ii) the system is bistable and appropriate choices of current injection are sufficient to switch between stable states; iii) synaptic efficacy parameters can be chosen such that current injection results in SWR-like events (similar to the scenario in Fig. 3.2).

Figure 3.6 shows that for an arbitrarily selected point, features i-iii are satisfied. Throughout the figure, injected currents are population-dependent and have been chosen to reduce transient effects upon stimulation. This means that the injected currents are sufficient for the system to jump to the SWR state, but are as low as possible to avoid overshoots. The synaptic efficacy values are $\tau_d = 250$ ms, $\eta = 0.3$. Figure 3.6A shows that, as expected, the 3d network (with fixed synaptic efficacy) has 3 fixed points (yellow crosses in upper plots). Note that the FP with non-zero coordinates (FP2) is visible in both B-P and A-P subspaces. Simulations with current injection confirm that the network (defined by Eq. (3.1)) is bistable when the synaptic efficacy is clamped at the value $e = 0.5$ (Fig. 3.6B) and exhibits SWR-like events when the synaptic efficacy evolves dynamically (Fig. 3.6C). This is true in all three conditions (figure columns), when current is injected to different populations. The simulations presented here use threshold-linear activation functions (Eqs. (3.4), (3.5), (3.6)), and only minor differences are visible when the biologically inspired softplus functions (Eq. (3.2)) are used. Thus, we can conclude that this parameter set represents a network that has all the required characteristics.

I predict that the results presented in Fig. 3.6 are similar for all points in the region defined by the derived conditions (green region in Fig. 3.5), and can verify this prediction by randomly selecting points and testing whether conditions i-iii are fulfilled. Limiting cases include parameter choices for which the system's nullclines intersect in three points, but ii and iii are difficult to verify. These are scenarios where the unstable FP is too close to one stable FP: this complicates the choice of the current to be injected for switching between the stable FPs. Another variable that might need to be tuned is the duration of the current pulse (the default duration is $T = 10$ ms to mimic optogenetic stimulation, see Sec. 3.6.2); indeed, in Fig. 3.6B, the pulse duration was set to $T = 50$ ms. This

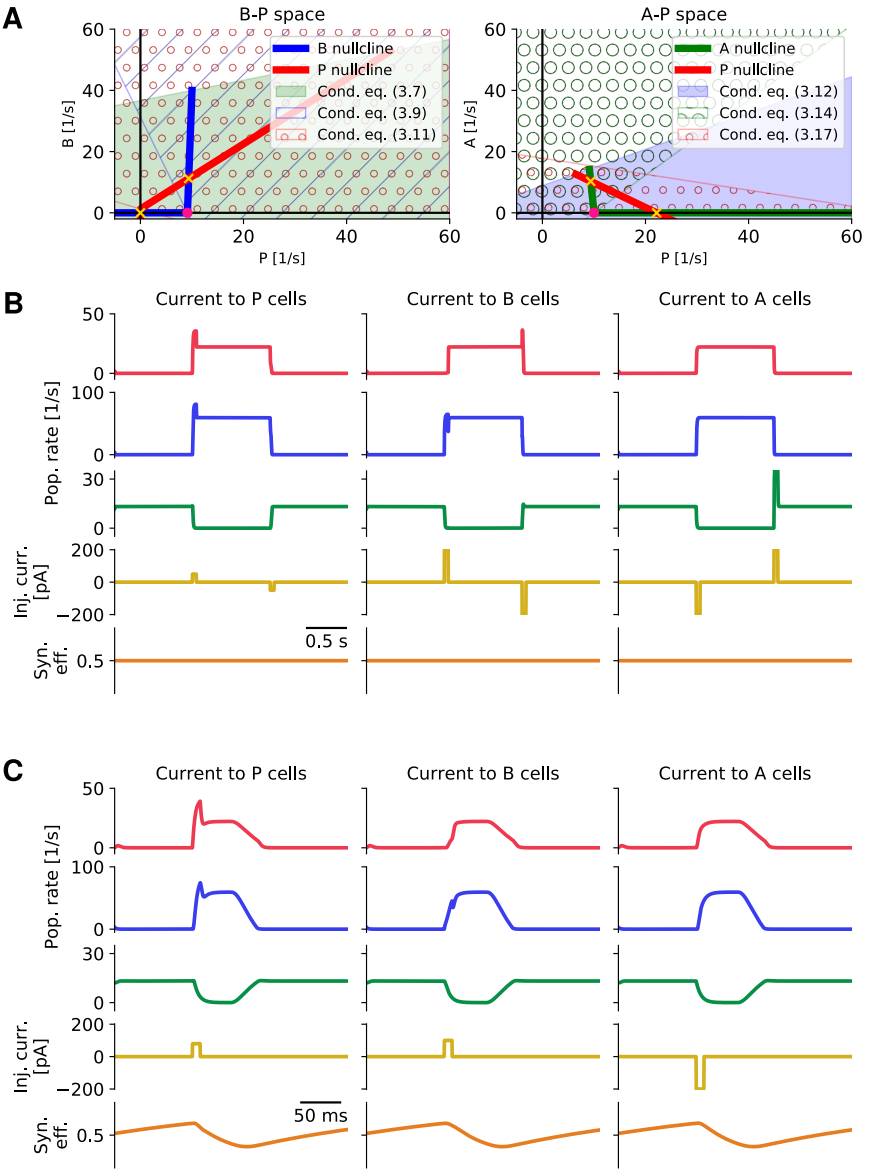

FIGURE 3.6: **Points inside the region defined by conditions 1-10 display bistability.** Simulation results are displayed for an example point in the green region of Fig. 3.5, with $W_{PP} = 2$, $W_{BP} = 8$, $W_{AP} = 1$, $W_{PB} = 2$, $W_{BB} = 3$, $W_{BA} = 13.5$, $W_{AB} = 6$, $W_{PA} = 10$, $W_{AA} = 5$, $k_P = 0.4$, $k_B = 0.5$, $k_A = 0.3$, $t_P = 130$, $t_B = 120$, $t_A = 110$ (and $\tau_P = 3$ ms, $\tau_B = 2$ ms, $\tau_A = 6$ ms, $\eta = 0.3$, $\tau_d = 250$ ms.). Note that this set of parameters cannot be visualized in Fig. 3.5 because more than two variables are changed with respect to the default network. **A)** Nullclines of the 3d system in the B-P space (left) and in the A-P space show two intersections in each panel, e.g. 3 FPs in total (the middle FP is visible in both plots). See Fig. 3.3 for plot details. **B)** Simulation results for the 4d network where synaptic efficacy is clamped to the value $e = 0.5$. Positive or negative currents of different intensities are applied to P, B, or A populations (left, middle and right plot, respectively) for $T = 50$ ms. Traces depict, from top to bottom: P, B, and A population rates, injected current, and synaptic efficacy. As expected, the network displays bistability and current injection is sufficient to switch between the two stable states. Firing rates in the non-SWR state are: $P = 0\,\mathrm{s}^{-1}$, $B = 0\,\mathrm{s}^{-1}$, $A = 13.2\,\mathrm{s}^{-1}$, and in the SWR state: $P = 22.2\,\mathrm{s}^{-1}$, $B = 59.5\,\mathrm{s}^{-1}$, $A = 0\,\mathrm{s}^{-1}$. **C)** Same structure as in B), but for a 4d network where the synaptic efficacy evolves dynamically as described by Eq. (3.1). Upon current stimulation, the network exhibits transient SWR-like events.

was necessary for the negative pulse to P and B to drive the system back to the resting (non-SWR) state, but was not needed in the case of A stimulation. This behavior hints at differences in the response properties of the populations depending on the stimulus type (positive vs. negative pulses) and on intrinsic properties of each population.

At this point, I stress again that the mean-field model I created exhibits SWR-like events upon current stimulation, but cannot generate spontaneous events because the system is deterministic. Possible extensions of the mean-field model could include a stochastic current component to elicit SWRs in the absence of stimuli. As the noise sources would be intrinsically different – network-driven in the spiking case, and external input in the rate scenario – it would be interesting to check in future work how the statistics of SWR-like events in the mean-field model compare to spontaneous events in the spiking network (see Chapter 2).

From the analysis presented here, we can conclude that the method I developed in Sec. 3.3 can be used to scan the parameter space and reliably identify regions where the 3d network [1] is bistable. Additionally, the current injection responses of the 4d network (with dynamic synaptic efficacy) resemble SWR events in experiments. This means that the effect of activating the P and B populations, or inactivating the A population, can be qualitatively similar in experiments, spiking model, and its mean-field approximation.

3.4.2 Outside of the region where conditions are fulfilled

By exploring the parameter space where all conditions presented in Sec. 3.3.1.4 are simultaneously fulfilled, I showed in the previous section that parameter sets in this region represent bistable mean-field networks, where the effect of current injection (positive current to P and B, negative current to A) is comparable to the results in the spiking network. This means that cases outside the green region in Fig. 3.5 are either not bistable, or they are bistable but the response to current injection does not match the results of the spiking model. To illustrate the latter scenario, I present in Fig. 3.7 an example of a point outside the green region of Fig. 3.5. The point has been chosen to display bistability (see Fig. 3.7A,B), while having $c < 0$ and $c' > 0$. As a consequence, conditions 1 and 2 in Sec. 3.3.1.4 (or equivalently, the requirements 3 and 5 defined in Sec. 2.3.1) are violated. The condition $c < 0$ means that the pathway $B \to \overset{\curvearrowleft}{A} \to P$ is weaker than the pathway $B \to P$. Thus, I expect that injecting positive current to the B population results in a decrease, rather than an increase, of P activity. Similarly, $c' > 0$ means that the pathway $A \to P$ is weaker than $A \to \overset{\curvearrowleft}{B} \to P$. Thus, the inactivation of A is expected to decrease P firing. In Fig. 3.7A, B, I show that the network is bistable, by looking at the nullclines in the B-P and A-P space (left and right, respectively) and at the effect of current stimulation when the synaptic efficacy is clamped at $e = 0.5$ (orange line). Note that in the left plot of Fig. 3.7B we observe the coexistence of the two states even if the negative current injection to P is not sufficient to bring the system back to the initial state (because in this example P has a larger value in the initial state). Additionally,

[1] or, equivalently, the 4d network with clamped synaptic efficacy

Figure 3.7C shows the effect of transient stimulation when the synaptic efficacy variable evolves dynamically. P activation (by positive current injection) transiently increases the firing rate of the P population before the high activity in B pushes it to a lower value. Similarly, current injection to B and A brings the system to a state where P activity is lower than the initial state. Note that the feature of the firing rate of the P population being lower in the SWR than in the non-SWR state is not a defining property of all points outside the green region in Fig. 3.5, but has been chosen for convenience in this example.

The results presented here confirm that the method I introduced goes beyond the identification of 3-dimensional bistable systems by accounting for specific requirements on the strength of the pathways in the network.

3.4.3 From mean-field models to new spiking models

The analysis of the nullclines allowed us to formulate a compact, quasi-analytic set of conditions (see Sec. 3.3.1.4) that combine the requirements on bistability with the requirements on the pathway strengths needed to build mean-field models matching experimental results. Despite the relative simplicity of these conditions, it is not trivial to understand how they could be used to tune mean-field networks to show the desired behavior (see Sec. 3.4.1). Moreover, can we use the knowledge derived from this analysis to predict instances of spiking networks with the desired properties? This section addresses this question, trying to link the mean-field models defined for parameter values inside the 'good' region (green region in Fig. 3.5) with spiking models matching experimental results.

However, the aforementioned question is ill-posed for several reasons. First, a spiking network has a much larger set of parameters than a mean-field network, which results in a non-univocal mapping from mean-field to spiking models. In other words, multiple spiking models can be derived from a single mean-field model. Second, as we have discussed in Sec. 2.6, the shapes of the f-I curves of the spiking model depend on multiple factors, including neuronal properties (i.e. reset and threshold voltage, refractory period, etc.), internal noise of the system and background injected current, which are quantities that are not defined in the mean-field model. Thus, it is unclear how to use the parameters of the rate activation functions (t_I, k_I, with $I \in \{P, B, A\}$, see Sec. 3.2) to constrain the properties of the f-I curves of the spiking network. This inevitably weakens the link between mean-field and spiking models. Third, as explained in Sec. 3.6.1, a drawback of using conductance-based synapses is the dependence of the W_{IJ} (average connection strengths from population J to population I in the mean-field model) on the mean steady-state membrane potentials of the populations. The issue is made even more complicated by the fact that we are interested in defining bistable systems where two stationary states (and two different voltage distributions) coexist. Thus, it is not clear how a given value of W_{IJ} can be related to parameters of spiking networks such as the

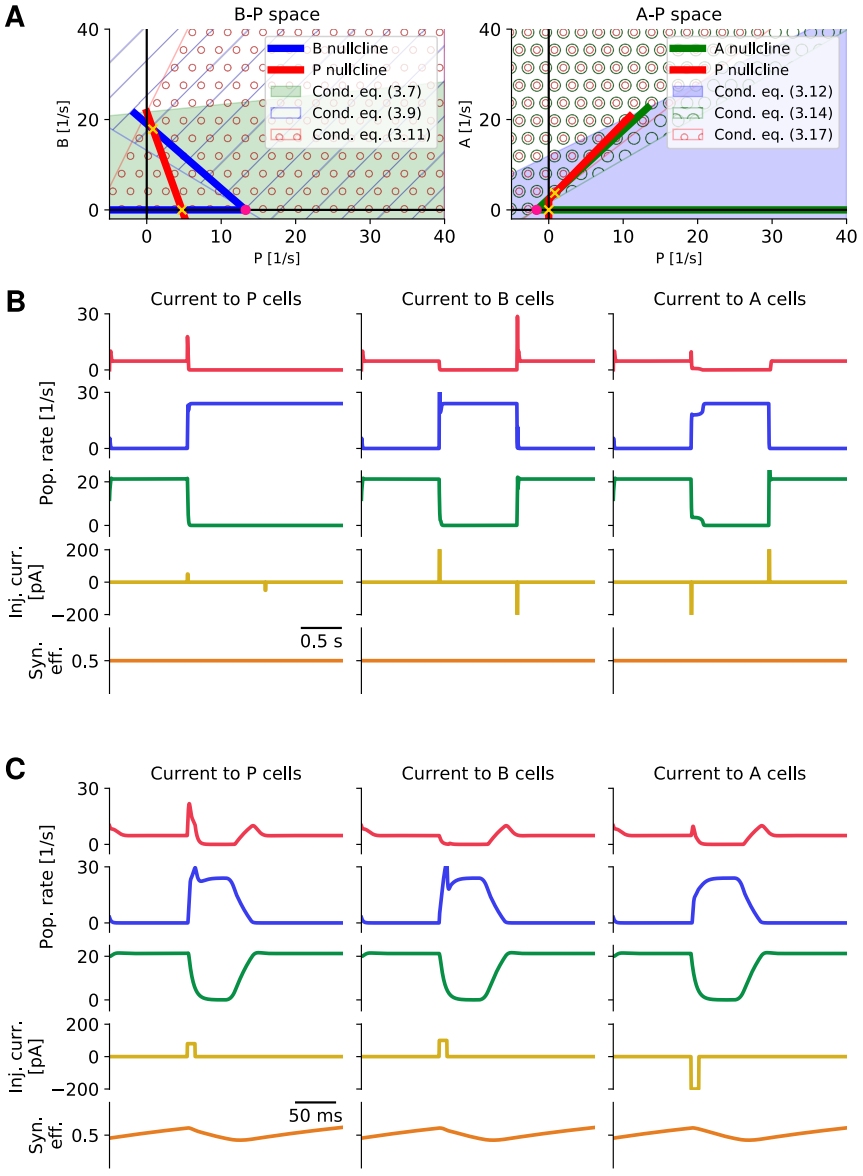

FIGURE 3.7: **Mean-field model with parameters outside of the green region in Fig. 3.5 can display bistability, but reversed pattern of activation.** Simulation results are displayed for a point outside of the green region of Fig. 3.5. The system displays bistability, but the current injection paradigm shows results that are different from the results presented in Fig. 3.6. This is due to the fact that conditions on pathway strength are violated (see text for details). The plot structure is identical to Fig. 3.6. **A)** shows nullclines in the B-P and A-P space (left and right plot, respectively). **B)** shows the coexistence of two bistable states when synaptic efficacy is clamped at $e = 0.5$. Current is injected for $T = 10$ ms. **C)**, from left to right, shows the effect of injecting current to P, B, and A populations. Note that injecting positive current to B and negative current to A results in a decrease of P activity. Current is injected for $T = 10$ ms. Note the scale difference with respect to panel B. Network parameters are: $\tau_B = 2$ ms, $\tau_P = 3$ ms, $\tau_A = 6$ ms, $t_P = 130$, $t_B = 120$, $t_A = 130$, $k_P = 0.4$, $k_B = 0.5$, $k_A = 0.3$, $W_{PP} = 2$, $W_{PA} = 6$, $W_{PB} = 6$, $W_{AP} = 1$, $W_{AA} = 3$, $W_{AB} = 6$, $W_{BP} = 8$, $W_{BB} = 3$, $W_{BA} = 10$, $\eta = 0.4$, $\tau_d = 300$ ms. Note that this set of parameters cannot be visualized in Fig. 3.5 because more than two variables are changed with respect to the default network.

connection probabilities p^{IJ} and the unitary conductance updates g_{ij}^{IJ} (see Sec. 3.6.4 for details).

Despite these limitations, I now show that mean-field models can guide the creation of spiking models. This example is meant to be a proof of principle showing that a mean-field model yields useful information on how to successfully build spiking networks as the one discussed in Chapter 2.

In what follows, I sketch a procedure by which new spiking models can be created. The method is explained in greater detail in Section 3.6.4. The starting point of this analysis is the arbitrarily chosen set of parameter values used to define the mean-field model presented in Fig. 3.6.

To be consistent with the spiking model presented in Chapter 2, I use leaky integrate-and-fire units with conductance-based synapses. To restrict the choice of possible spiking models that can be defined from the mean-field model, I make the simplifying assumption that all neuronal parameters (voltage reset and threshold, refractory periods, membrane conductance, etc.) are fixed, and correspond to the standard values used in spiking networks (see Sec. 2.8.1 and Table 2.1). Next, I have to assume a connectivity structure in the network, to be able to map the connection strengths W_{IJ} to the corresponding spiking-model parameters, i.e, the connection probability between populations p^{IJ} and the unitary conductance update g_{ij}^{IJ}. For simplicity, I fix all the connection probabilities in the network to $p = 0.6$ and later discuss how this choice impacts the results. As previously discussed, the choice of conductance-based synapses complicates the transition from W_{IJ} to spiking neurons' connections because the effectiveness of an incoming spike depends on the driving force (reversal potential minus the average membrane potential of the presynaptic population, see Eq. (3.72)). To overcome this issue, I run an optimization procedure on the membrane potential values V_P, V_B, and V_A and select the values for which the spiking model best matches the mean-field population rates (see Fig. 3.6). The optimization procedure is analogous to what I have introduced in Sec. 3.2 and is described in more detail in Sec. 3.6.4. The optimal values are, for the mean-field model shown in Fig. 3.6, $V_P = -55$ mV, $V_B = -54$ mV, and $V_A = -58$ mV. These values can be used to define the unitary conductance updates g_{ij}^{IJ} for each synapse type (see Sec. 3.6.4 and Table 3.2).

In Fig. 3.8, I show the dynamics of the resulting spiking model. Panel A shows that the system exhibits bistability when the synaptic efficacy is clamped at an intermediate value (here, $e = 0.5$), and current injection can trigger a switch between the states. The population firing rates in the spiking model show a good quantitative match with the population rates of the mean-field model (maximal distance between population firing rates is $\sim$2 s^{-1}, compare this figure with central panel in Fig. 3.6). In the SWR state, P and B cells display large synchronicity of firing that results in large population oscillations; this is likely due to the large connection probabilities ($p = 0.6$ at all connections, see Table 2.2 for a comparison with the more biologically plausible network presented in Fig. 2.11, Chapter 2). The other panels of the figure (Fig. 3.8 B, C, D) show that the 4d network, with dynamic synaptic efficacy, shows SWR-like events upon positive current

injection to P and B, and negative current injection to A. Synaptic efficacy parameters for the connection $B \to A$ ($\tau_D = 300$ ms, $\eta_D = 0.4$) have been chosen to yield events with a duration close to experimentally observed values (50-100 ms).

As discussed above, different spiking networks can be created by choosing a different connectivity scheme (i.e., combinations of connection probabilities and unitary conductance updates). In general, these networks will still yield results similar to those displayed in Fig. 3.8, even though the results are expected to vary in terms of quantitative match to the mean-field model and synchrony levels in both stationary states. The match with the mean-field model cannot be achieved for too extreme choices, as connectivity values larger than $p = 0.8$ (in this case, the input variability across cells is too low and the populations are synchronized), or connections that are too sparse (e.g., if $p < 0.6$ for connections starting from A cells, the postsynaptic neurons receive on average too few connections, see Sec. 2.8.2).

It is now interesting to check how the optimized mean membrane potentials relate to the simulated values of the populations in the two stationary states. How far off are the optimized choices of the mean membrane potentials V_I (used to link W_{IJ} to spiking-model parameters) from actual mean membrane potentials of the spiking model? Figure 3.9 shows that, as expected, in the bistable spiking network the mean membrane potential values change depending on the stationary state the network is in (as discussed in Sec. 3.6.1). As a result, the distributions are bi-modal. The optimized values of $V_P = -55$ mV, $V_B = -56$ mV, and $V_A = -58$ mV (black lines in the probability density plots) provide a compromise between the values in the two network states for P and B cells. For the A population, however, the optimal value is mostly driven by the membrane potential in the SWR state.

As a last control on the correspondence between the two models, I retrospectively analyze whether the activation functions of the mean-field model are good predictors of the shapes of the spiking f-I curves. Figure 3.10 displays the f-I curves for each population (colored lines, with P, B, and A organized from left to right) in each stationary state (A): non-SWR state, (B): SWR state. The black lines are the threshold linear functions (see Eqs. (3.4), (3.5), (3.6)) of the mean-field model described in Fig. 3.6, from which the spiking model was derived. The curves should match in the shaded areas in the figure, which indicate the probability that an input current at that current amplitude arrives to each population in either state. Overall, we observe that the mean-field model well predicts the firing behavior of P and B cells in both states. However, the input current in the mean-field model is overestimated for the A cells, with respect to the spiking network. This is likely due to the fact that in the mean-field model, both P and B are zero in the non-SWR state, whereas they are close to 1 s^{-1} in the spiking model. Because the connection $B \to A$ is expected to be strong, whereas the connection $P \to A$ should be weak (see Sec. 2.3.1), the net effect of non-zero firing of P and B cells in the spiking model's non-SWR state is inhibitory, and the input level is thus shifted to lower values. Similarly, in the SWR state, the non-zero firing of A cells reduces the net input current received by the A cells, compared to the mean-field scenario where $A = 0$. The

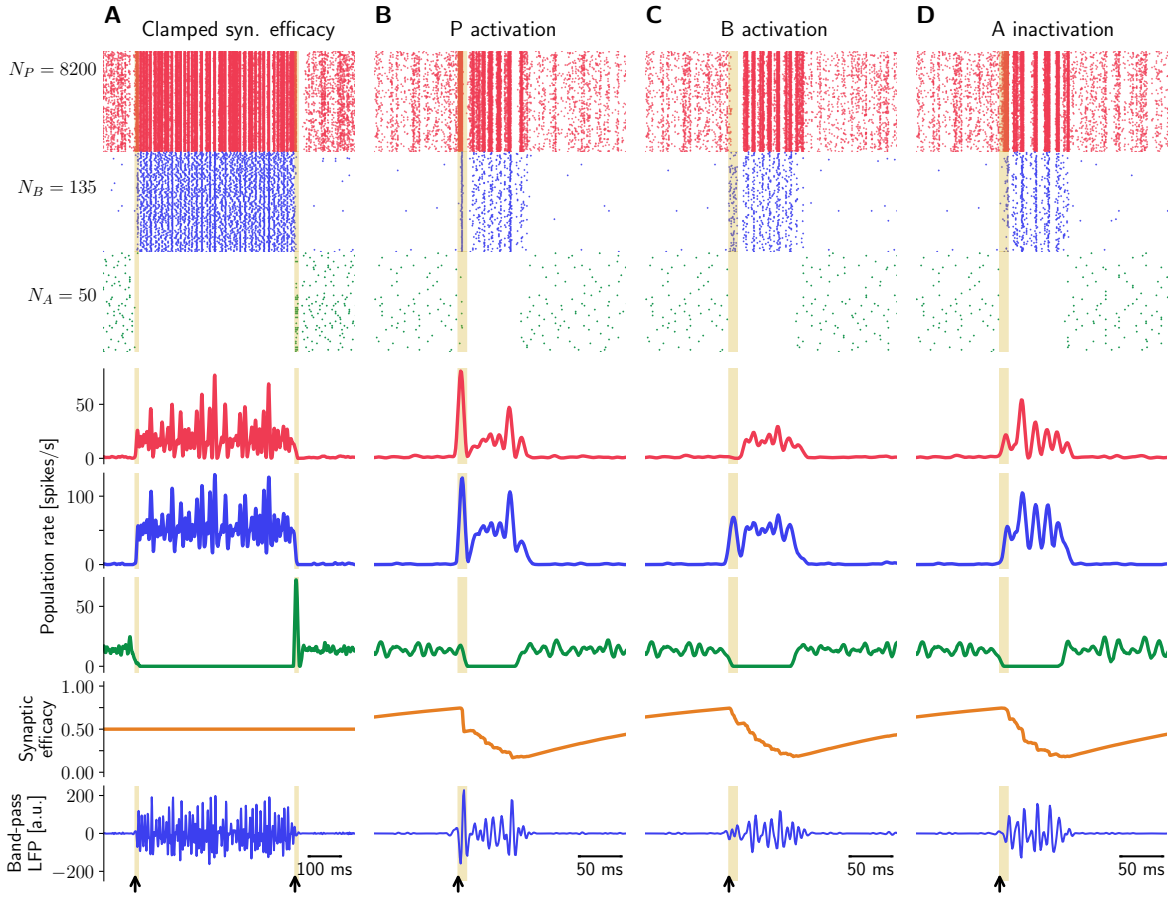

FIGURE 3.8: **Example of a spiking model that displays bistability and SWR-like behavior**. Displayed are simulations for a spiking network derived from the mean-field model presented in Fig. 3.6. Connection probability was chosen to be $p = 0.6$ for all synapses; to define the conductance update values, the optimized values $V_P = -55$ mV, $V_B = -54$ mV, $V_A = -58$ mV, are used (see text for details). Traces represent, from top to bottom: raster plots for P, B, and A populations, instantaneous population rates for P, B, and A cells, mean synaptic efficacy value at synapses $B \to A$, and band-pass (90-180 Hz) filtered component of the LFP. The plot structure is the same as Fig. 2.2. **A)** Bistability in the network with clamped synaptic efficacy. Upon negative current injection to A cells, the system switches from the non-SWR to the SWR state. Positive current injection to A cells brings the system back to the non-SWR state. Average firing rates in the non-SWR state are: $P = 1.4$ s^{-1}, $B = 0.1$ s^{-1}, $A = 13.6$ s^{-1}, and in the SWR state: $P = 21.6$ s^{-1}, $B = 59.5$ s^{-1}, $A = 0$ s^{-1} (compare this with firing rates of the mean-field model, Fig. 3.6). **B - D)** SWR-like events emerge in the network, induced by positive current injection to P and B cells (panels B and C) and negative current injected to A cells (panel D). Synaptic efficacy parameters are $\tau_D = 300$ ms, $\eta_D = 0.4$. In all panels, current is injected for $T = 10$ ms (injection periods are marked by yellow shaded areas); current amplitude is varied uniformly across cells (see Sec. 3.6.4).

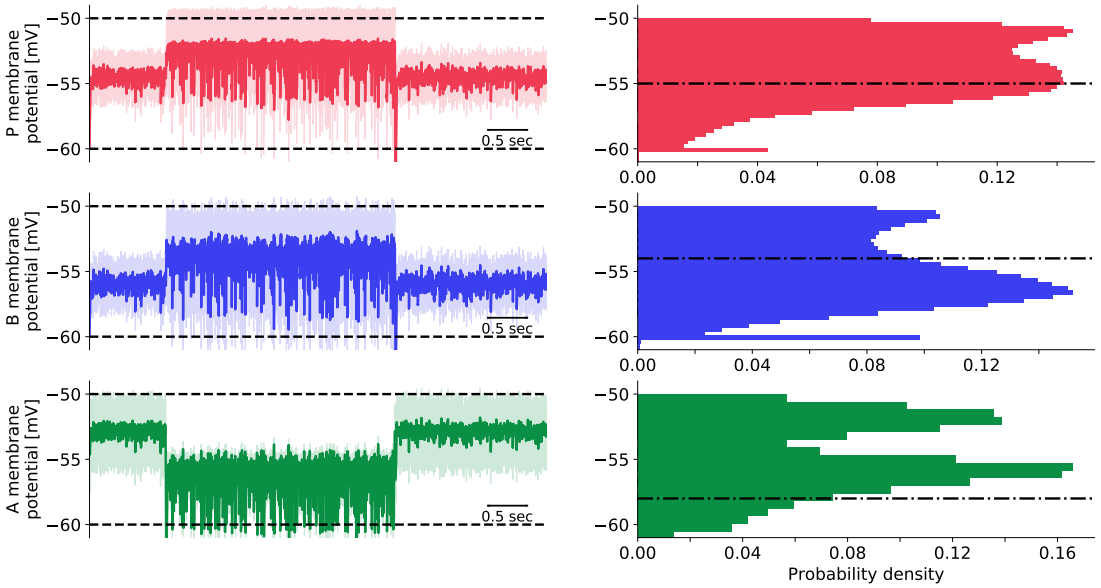

FIGURE 3.9: **Membrane potential distributions of the bistable spiking network.** The spiking network presented in Fig. 3.8 is simulated for 6 seconds, and the membrane voltages of all cells in each population are monitored. Negative current is injected to the A cells at $T = 1$ second to let the system jump from the non-SWR state to the SWR state, and positive current is injected to the A cells at $T = 4$ seconds to jump back to the original state. Left: mean membrane potentials over time (solid lines) $\pm$ one standard deviation (shaded area). Statistics are calculated across all cells in a population. From top to bottom: membrane potential of P, B, and A cells. Black dashed lines represent the membrane threshold and reset values. Right: histograms of membrane potentials shown in the left plots. The black lines indicate the vales obtained from the optimization procedure ($V_P = -55$ mV, $V_B = -54$ mV, and $V_A = -58$ mV).

comparison shows the limitations of the approach presented here and reminds us that the two models can be constructed to have similar firing rates, but do not share the same properties.

3.5 Discussion

In this chapter, I have defined a mean-field approximation of the spiking model presented in Chapter 2, and I have shown that this model can be used to understand the mechanisms responsible for the existence of networks that are both bistable and respond with the correct activation pattern upon current injection.

A general limitation of this approach is that the mean-field networks are deterministic, i.e., spontaneous SWR-like events cannot be generated. This is a consequence of the bistable structure of the network with intermediate synaptic efficacy values (e.g. $e = 0.5$). Spontaneously occurring SWR events could be obtained by inserting a noise component in the input to the mean-field model. Another possibility could be to study regimes where the network is oscillating, rather than being in a bistable configuration, such that SWR events would emerge periodically in the network. It remains to be clarified whether this approach could yield short-duration SWRs interleaved with long periods of silence,

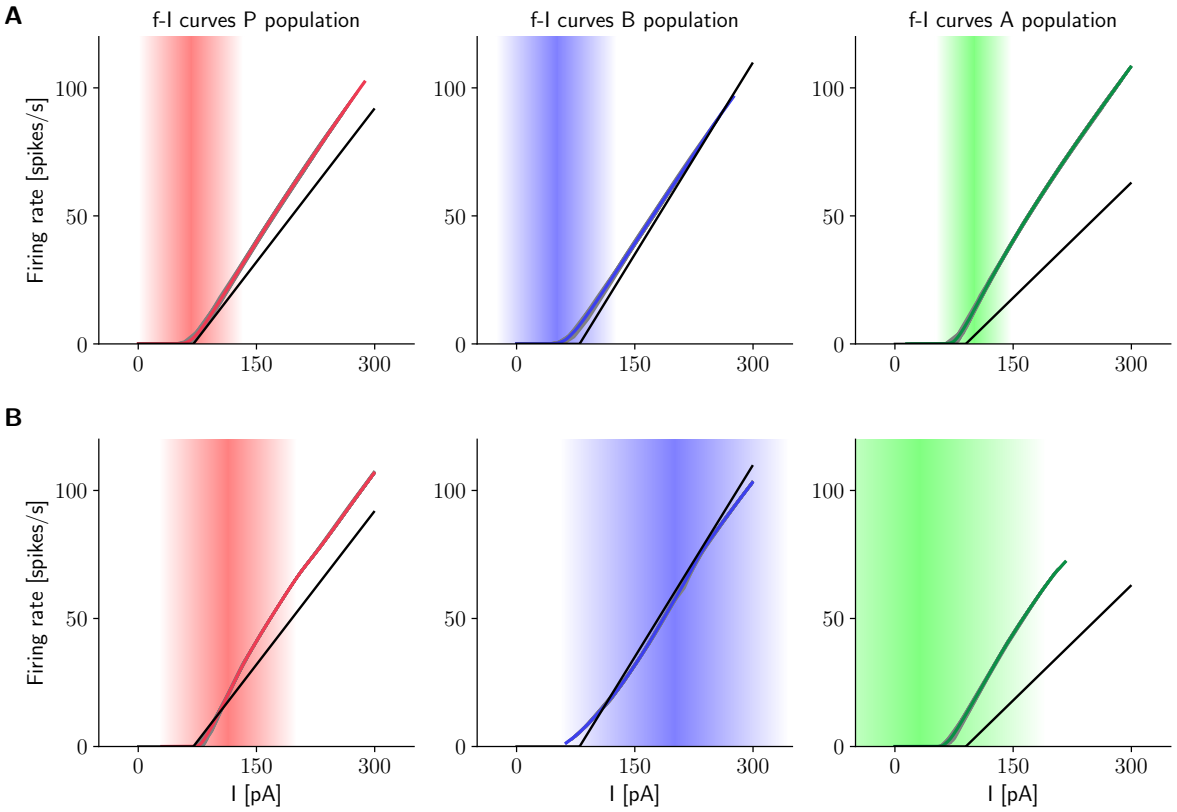

FIGURE 3.10: **Mean-field activation functions do not accurately match f-I curves of the new spiking network. A)** f-I curves of spiking populations in the non-SWR state (P, B, and A cells, from left to right), for the network displayed in Fig. 3.8. Single neuron f-I curves are plotted in gray. Colored curves are the mean f-I curves, averaged across 50 neurons in each population. The shaded areas depict the values of the mean total input current ($\pm$ one standard deviation). The black curves depict the activation functions of the mean-field model displayed in Fig. 3.6, from which the spiking model has been derived. **B)** Same as A), but for the network in the SWR state. The plots show that the activation functions of the mean-field model are able to partially predict the populations firing behavior in the spiking network. Note the good match for P and B populations in the shaded areas in both states, and the mismatch in the case of A cells in the SWR and non-SWR states.

as shown by experimental results (see Sec. 1.2.3). Also, to mimic the variable occurrence of SWRs reported in experiments, noise should be added to the input of the oscillating mean-field model to alter its clock-like structure.

In Sec. 3.2 I have presented a method to derive a mean-field network from a conductance-based LIF network where two stationary states coexist. I have shown that the mean-field model can quantitatively match the stationary firing rates of the spiking model. However, a major drawback of the approach is that the mean membrane potentials of the three populations need to be fitted in order to compute the connection strengths in the mean-field model. The choice of optimizing those values might seem at first arbitrary, but it is supported by the results showing that the optimal values lie in between the simulated mean membrane potentials in either state. From a biological perspective, it is

not surprising that two different network states are characterized by different mean membrane potentials, so more effort should be oriented to extending the theory to multistable configurations.

More theoretically grounded methods have been developed to treat bistability in the context of spiking networks approximations, mostly in relation to UP-DOWN states modeling [Renart et al., 2003; Montbrió et al., 2015; Schwalger et al., 2017; Di Volo et al., 2019]. However, in these approaches the transfer functions of the populations cannot be derived explicitly. Importantly, I made use of the existence of an explicit formula for the transfer functions in Sec. 3.3, to derive bistability conditions for the network.

An alternative strategy to link spiking and mean-field models could be to start from a spiking network with current-based synapses. This approach would reduce the biological realism of the spiking model, but could simplify the derivation of the mean-field model, as more literature is available on the approximation of current-based LIF neurons (see e.g. [Renart et al., 2003]). In the approach presented here, current-based synapses would remove the dependence of the connection strength in the mean-field model on the mean membrane potentials, thus making the link between spiking and mean-field model more direct.

The main result of this chapter is the definition of bistability conditions in the mean-field model. The method is based on the explicit computation of nullclines and fixed points, and is fully analytic (except for the instability condition on the middle FP). To simplify the analysis, I assumed that the B and A populations are zero in the non-SWR and SWR states, respectively. However, this assumption can be relaxed in order to extend the analysis to networks where B and A are constant (but not necessarily zero) in their inactive states. Additionally, I expect that the analysis could be easily extended to study cases where the activation functions are piecewise-linear, or more generally such that the inverse function can be explicitly derived. Thus, I have constructed a general framework that can be applied to different types of 3-dimensional networks. The set of conditions presented in Sec. 3.6.3.4 defines a region in parameter space where a mean-field network can be created, that fulfills the experimental results on SWR initiation by cell stimulation discussed in Sec. 1.2.3. A main achievement of the analysis consists in the definition of clear criteria to understand the contribution of each variable to the existence of mean-field networks showing the desired behavior. However, due to the large number of parameters and conditions, it is not easy to visualize the region of bistability, nor to predict the effect of the perturbation of a parameter on the network behavior. Additional work should be devoted to develop a more compact formulation of the conditions and a method to efficiently visualize the effect of different parameters.

How can the knowledge derived in the mean-field network be used to create a spiking model? In Sec. 3.4.3, I presented an example of a spiking network which was defined starting from a mean-field network. In doing so, I necessarily had to restrict the set

of parameters of the spiking model. I chose to use the same neuronal parameters as the spiking network presented in Chapter 2 (but a different connectivity scheme) and showed that the spiking model displays features similar to the mean-field model. Future work should address how modifying neuronal parameters affects the feasibility of this approach.

While creating the spiking model, I had to resort to an optimization procedure to link the mean-field model parameters with connectivity parameters. As discussed earlier in this section, a current-based approach could simplify the link between the two models. Alternatively, the match between spiking and mean-field models could be refined by iteratively creating spiking and mean-field models in sequence. The knowledge about how to approximate spiking models using mean-field models could be used to create a better approximation of the newly created spiking model, and the spiking model itself could be improved using the theory developed in Sec. 3.4.3. In this way, the distance between the two models, in terms of population firing rates, could possibly be minimized. It remains however to be tested how this method would deal with the existence of two stationary states. Creating a closer correspondence of other network properties requires the development of a more involved theory, e.g., to match the noise structure in the two networks or to understand how neuronal properties influence the shape of the f-I curves.

Another issue that makes the correspondence between the mean-field and the spiking model difficult is the missing link between the activation functions in the two models. As shown in Fig. 3.10, the activation functions of the mean-field model do not match the spiking network f-I curves well. The analysis of this discrepancy goes beyond the scope of this work, but the issue should be taken into account in future work.

More generally, I used the knowledge derived in Sec. 3.3 to define a mean-field model and used this as a starting point to construct a new spiking model. This heuristic method is meant to be a proof of principle showing that the method can work, but additional work should address the robustness of the method in a number of conditions (e.g., depending on the starting mean-field network, the choice of connectivity, etc.).

An important aspect that should be remembered when creating new spiking models is that the theory I have developed can at most generate spiking networks that are bistable and mimic experimental results on the effect of current injection. However, the method does not guarantee the existence of spontaneous SWR events in the new spiking network. As discussed in Sec. 2.8.3, spontaneous events (and their dynamics) depend on the complex interplay of the synaptic efficacy parameters with the strength of the reciprocal connections among interneurons. Thus, it is non-trivial to understand how to modify a spiking network obtained from a mean-field model to yield spontaneous events. This aspect, already discussed in Sec. 2.7, constitutes an important limitation of the method.

3.6 Methods

3.6.1 Definition of mean-field parameters

Activation functions. I use the stationary single neurons' f-I curves to define the activation functions $f_I(x)$, for $I \in \{P, B, A\}$ (Eqs. (3.4), (3.5), (3.6)). As the network is bistable for intermediate values of synaptic efficacy (i.e. when $e_{IJ} = 0.5$, see Sec. 2.6), two sets of f-I curves need to be considered for each population. I present here the procedure to derive $f_B(x)$ and describe how to generalize it to the other populations. In general, $f_B(x)$ should describe accurately the input-output relation for lower input levels in the non-SWR state, and for higher input levels in the SWR state. Indeed, Fig. 3.11 shows that the mean input to B is around 70 pA in the non-SWR state, and it is close to 300 pA in the SWR state. Thus, I define an empirical f-I curve by taking the mean f-I curve of the spiking network in the non-SWR state below a given threshold, and the mean f-I curve of the SWR state above the threshold. The threshold is defined as the current value where the mean input current minus one standard deviation arrives to the B population in the SWR state. This state can be considered as the 'active' state for B cells because they are almost silent in the non-SWR state. I then fit this empirical f-I curve to a softplus function $f(x) = \ln\{1 + \exp[k(x + s)]\}$ [Dugas et al., 2001; Glorot et al., 2011], where the parameters k and s are optimized via least-square error minimization. The softplus function shows a convex increase for small x and grows linearly as $g(x) = k(x+s)$ for large x. For the fitting, k is constrained to the interval $[0, 2]$ and s to the interval $[-100, 0]$. Optimal values for $f_B(x)$ are $k_B = 0.41$, $s_B = -68.04$.

Because the 'active' state of the P population is the SWR state, the exact same procedure described above applies to the f-I curves of P. In the case of the A population, whose 'active' state is the non-SWR state, the only difference is that the empirical f-I curve is defined by considering the mean f-I curve of the SWR state below threshold, and the mean f-I curve of the non-SWR state above threshold. For the simulations in Figs. 3.1 and 3.2, I use the fitted softplus functions (with $k_P = 0.47$, $s_P = -68.34$, $k_A = 0.48$, $s_A = -68.91$), for which I additionally define $t_I = s_I + 200$ as explained in the main text.

To further simplify the activation functions so that they are amenable to an analytical treatment, in Section 3.3 I approximate them using threshold linear functions defined as

$$f_I(x) = \begin{cases} k_I(x + t_I) & \text{if } x \geq -t_I \\ 0 & \text{else} \end{cases} \qquad I \in \{P, B, A\} \tag{3.21}$$

In doing so, we are effectively setting the firing rate of the populations in their 'inactive' state to 0 s^{-1}. This choice is motivated by the fact that the firing rates are close to zero in the spiking model (see Fig. 2.11), and that it enables the analysis of fixed point stability.

Connections W_{IJ}. The average synaptic strength between two populations should depend on the size N_I of the presynaptic population, the connection probability p^{IJ}, the

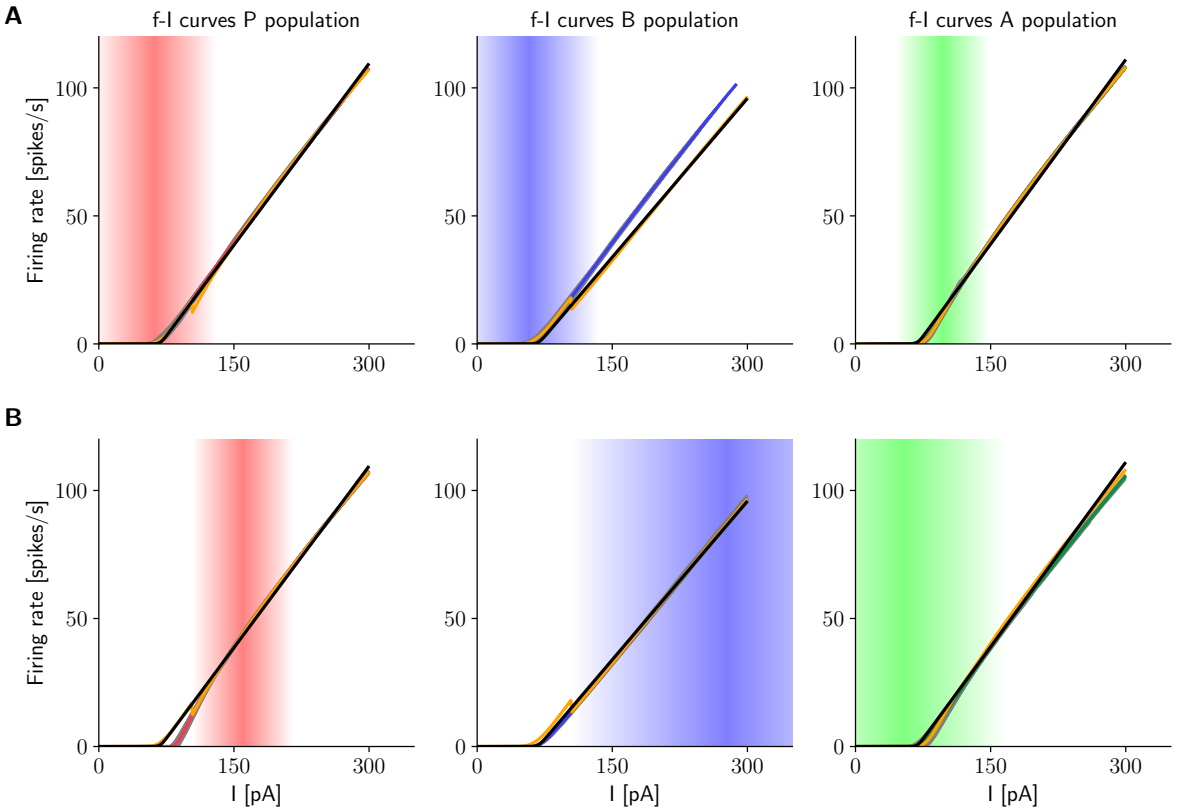

FIGURE 3.11: **Definition of mean-field activation functions.** Spiking f-I curves are defined as in Fig. 2.11. **A)** non-SWR state; **B)** SWR state. In both panels, f-I curves are shown, from left to right, for P, B, and A cells. In addition: Orange curves define the weighted f-I curves (one per population, see text for details). Black curves are the best fit softplus functions (one per population). The fitted curves are the activation functions used in the simulations shown in Figs. 3.1 and 3.2.

unitary conductance updates g^{IJ} when a presynaptic spike occurs, the synaptic drive and the conductance decay time constants τ^P_{syn}, τ^B_{syn}, and τ^A_{syn} [Gerstner et al., 2014]. More formally, we can define the W_{IJ} as

$$\begin{aligned} W_{IP} &= N_P p^{IP} g^{IP}_{ij} \tau^P_{syn}(E^P_{rev} - V_I), \quad I \in \{P, B, A\} \\ W_{IB} &= -N_B p^{IB} g^{IB}_{ij} \tau^B_{syn}(E^B_{rev} - V_I), \quad I \in \{P, B, A\} \\ W_{IA} &= -N_A p^{IA} g^{IA}_{ij} \tau^A_{syn}(E^A_{rev} - V_I), \quad I \in \{P, B, A\} \end{aligned} \tag{3.22}$$

Note that, for simplicity, I neglect the synaptic delays in this approximation. The terms V_I should describe the average membrane potential values of cells in the postsynaptic population I. However, in our bistable scenario, the average population membrane potentials differ across the two stable states (because the inputs each cell is receiving change across states, as shown in Fig. 3.12). Thus, there is no predetermined way of defining these values. However, many combinations of V_P, V_B, and V_A result in mean-field models with biologically-realistic firing rates, as displayed in Fig. 3.13. For this reason, I decided to keep V_P, V_B, and V_A as free parameters and run an optimization procedure that searches for values that minimize the distance between the target population firing rates

	P [s^{-1}]	B [s^{-1}]	A [s^{-1}]
non-SWR state	< 5	< 5	> 8
SWR state	> 8	> 30	< 5

TABLE 3.1: Summary table for 'biological' population firing rates in non-SWR and SWR states.

in the spiking model (see Fig. 2.11) and the population rates of the mean-field model. In both cases, the synaptic efficacy is clamped to ensure bistability. More in detail, V_I ($I \in \{P, B, A\}$) can range from the reset to the threshold potential. For each possible combination of V_I in this range ($[-60, -50]$ mV, using a step size of 0.5 mV), I run a mean-field simulation for $e = 0.5$ (clamped synaptic efficacy), using the fitted softplus activation functions, and $\tau_P = 3$ ms, $\tau_B = 2$ ms, and $\tau_A = 6$ ms as described in the main text. The system is initialized to start from the non-SWR state. Current is injected to the P and B populations (positive current) and to the A population (negative current) to let the system jump to the SWR state. I store the population values in both states if the stimulation is successful, i.e., *i)* the same two stable states coexist in all three stimulation paradigms, ii) the firing rate of the stable states are comprised in a 'biological' range (close to experimental results, see Table 3.1). Finally, I minimize the Euclidean norm between the vector of target firing rate values in the spiking model and the vector of firing rates in the mean-field model to find the optimal combination of V_I. Note that, in this way, the firing rates in the 'active' state of each population (SWR state for P, B, non-SWR state for A) are better matched than the ones for the 'inactive' state, which are close to zero. This is a reasonable choice, as the 'active' states are the ones that better characterize the firing of a population.

For the network configuration presented here, the optimized value are $V_P = -52.5$ mV, $V_B = -54$ mV, and $V_A = -52.5$ mV (also see Fig. 3.13). For B and A, these values are close to the mean membrane potential values in the 'active' state (Fig. 3.12). For the P population, the optimal value is an average of the peaks of the distributions of membrane potentials in the two states. This suggests that the optimization yields meaningful results. I use the optimal values of V_I to define the connections W_{IJ} as described by Eq. (3.22), and use these values to define the mean-field model used for simulations shown in Figs. 3.1, 3.2.

3.6.2 Mean-field model simulations

For the simulations presented in Sec. 3.2.2, networks are initialized to start close to the non-SWR state. In Fig. 3.1, current is injected to the P population ($I = \pm 100$ pA) for a duration of $T = 10$ ms. Similarly, in Fig. 3.2, each population is stimulated with a current pulse lasting $T = 10$ ms. The amount of injected current is chosen to minimize the transient effects, and correspond to $I_P = 60$ pA, $I_B = 150$ pA, and $I_A = 200$ pA.

For the mean-field network shown in Figs. 3.6 and 3.7, the same approach is used. The only difference between the figures are the parameter choice (see figure captions)

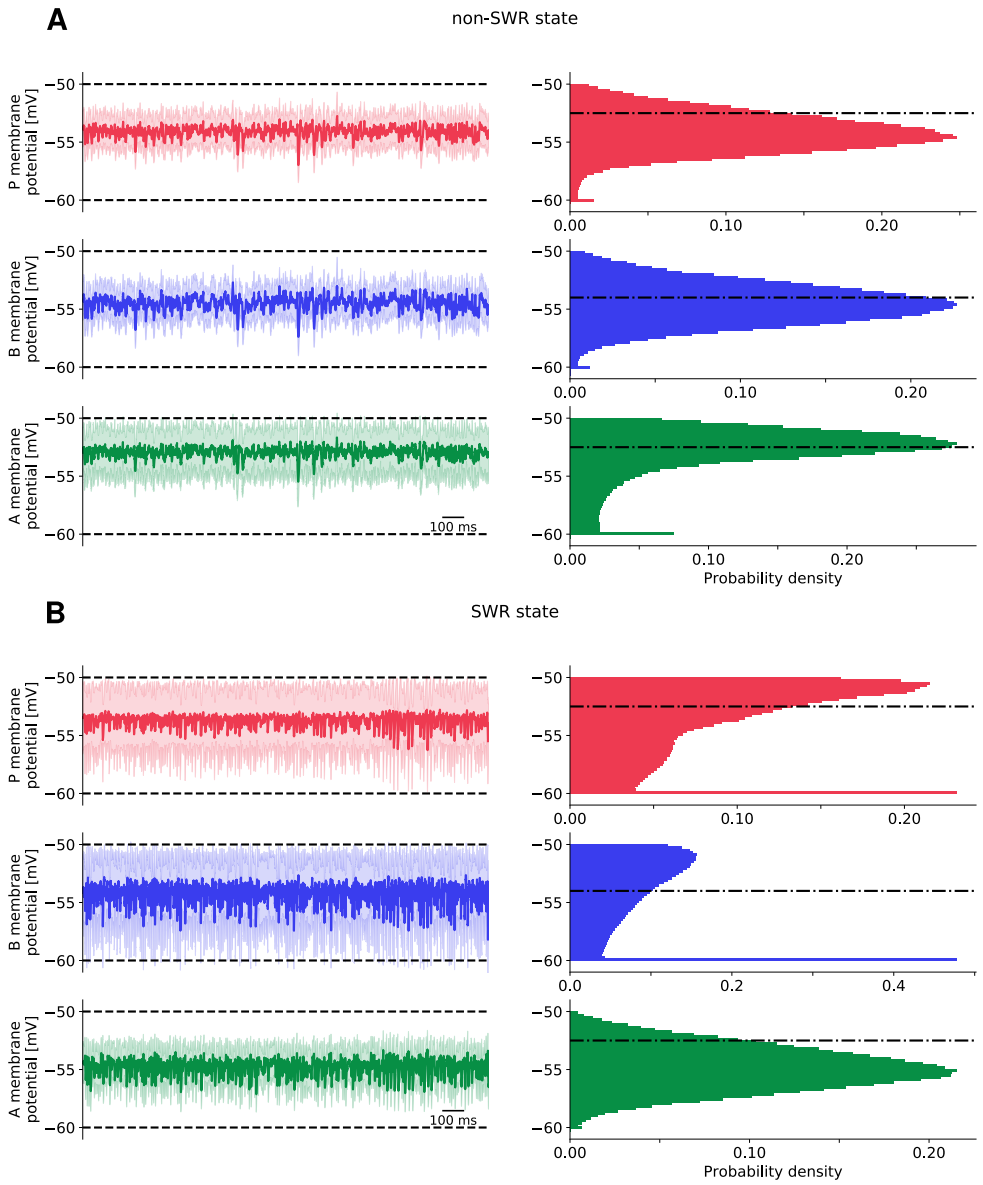

FIGURE 3.12: **Membrane potential values in spiking bistable network.** The spiking network presented in Fig. 2.11 is simulated for 2 seconds, and the membrane voltages of all cells in each population are monitored. The synaptic efficacy is clamped at $e_{IJ} = 0.5$ to ensure bistability. **A)** Left: mean membrane potentials over time (solid lines) $\pm$ one standard deviation (shaded area) in the non-SWR state. Statistics are calculated across all cells in a population. From top to bottom: membrane potential of P, B, and A cells. Black dashed lines represent the membrane threshold and reset values. Right: histograms of membrane potentials shown in the left plots. The black lines indicate the values obtained from the optimization procedure ($V_P = -52.5$ mV, $V_B = -54$ mV, $V_A = -52.5$ mV). **B)** Same as A), but for the SWR state.

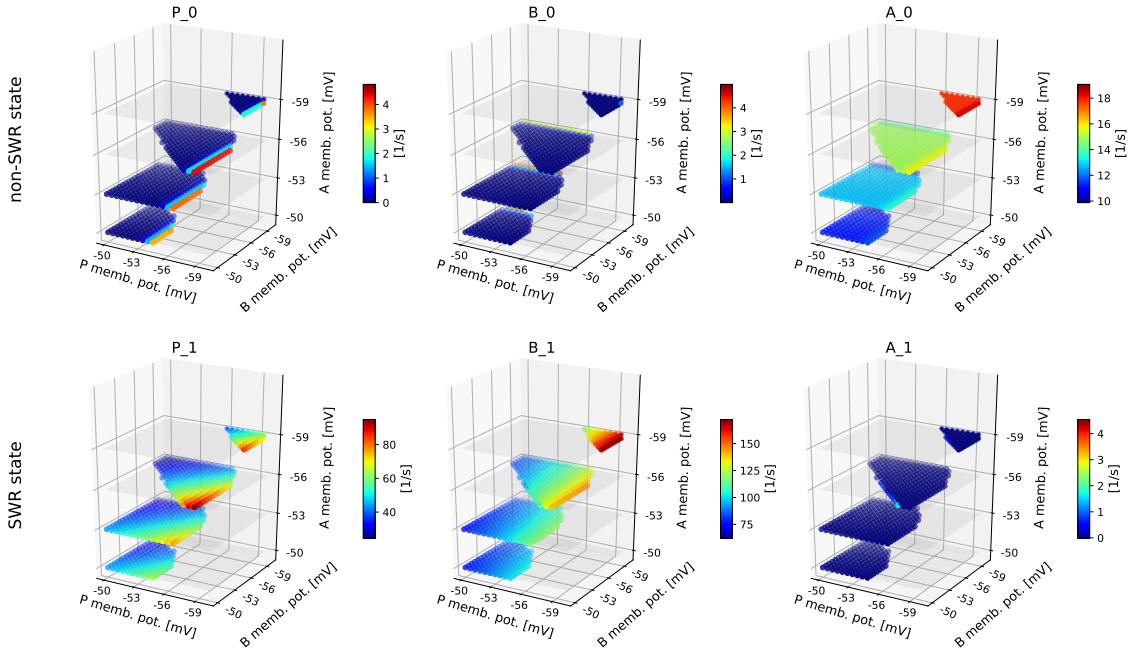

FIGURE 3.13: **Mean-field population rates as a function of $\mathbf{V_I}$.** Each column shows the mean-field stationary firing rates for P, B, and A population in non-SWR state (top) and SWR state (bottom). The colors code for the firing rate as the values of the mean membrane potentials are changed. Only firing rate values within 'biological' ranges (to match experimental firing rates, see Table 3.1 and Section 1.2.3) are shown. Most of the choices for V_I yield reasonable results even if the actual firing rate values can vary greatly depending on V_I. The optimal combination of mean membrane potentials is found by minimizing the distance between spiking and mean-field firing rates, and is, for the spiking network presented in Fig. 2.11, $V_P = -52.5$ mV, $V_B = -54$ mV, and $V_A = -52.5$ mV.

and the fact that, in Fig. 3.6B, current was injected for $T = 50$ ms instead of the default value of $T = 10$ ms. This was needed for the system to jump down to the non-SWR state when negative current was injected to the P and B populations.

3.6.3 Analytical results on network bistability

In this part, I present the details on the analysis of the 3d system discussed in Sec. 3.3. Starting from the network defined in Eq. (3.3), I derive equations for the nullclines and fixed points of the system. This allows us to define parameter regions where the mean-field model is bistable and can generate SWR-like events upon activation of different populations, mimicking experimental results (see Sec. 1.2.3).

In what follows, I look at the steady-state behavior of the system setting the left hand side of all equations in Eq. (3.3) to zero and calculating the nullclines. I aim at computing under which conditions the system shows three fixed points (FPs), two stable and one unstable. One stable FP should correspond to the non-SWR state, with A population active, P population only mildly active, and B population almost inactive (FP0 in what follows). The other stable FP is the SWR state, with P and B strongly

active and A almost inactive (FP1 in what follows). The unstable FP corresponds to a middle FP between FP1 and FP0, which behaves like a threshold for the system to switch between the two states (and is called FP2). All FPs define firing rates and should be non-negative. To simplify the analysis, we will assume that $B = 0$ s^{-1} in FP0 and $A = 0$ s^{-1} in FP1.

3.6.3.1 The B-P space

To derive an equation for the nullclines, we first restrict ourselves to the 2d space B-P, assuming that $A > 0$ (i.e. we exclude for now the case in which $A = 0$). We set $\frac{\partial A}{\partial t} = 0$ in Eq. (3.3c), solve this equation for A and use it to compute the P and B nullclines in this space. The nullclines should intersect in two points, corresponding to the non-SWR fixed point (FP0) and the middle FP (FP2). Both FPs have a non-zero A coordinate. Note that under these assumptions, we cannot see the SWR FP (FP1) because we have assumed that its A coordinate is 0.

Assuming that $A > 0$ means that the input to f_A (defined in Eq. (3.6)) is larger than the threshold. Thus, we assume that

$$W_{AP}P - W_{AB}B - W_{AA}A \geq -t_A. \tag{3.23}$$

If this is the case, we can solve Eq. (3.3c) for A (in the steady-state, with non-zero branch of Eq. (3.6)) and obtain

$$A = \frac{k_A}{1 + k_A W_{AA}}(W_{AP}P - W_{AB}B + t_A). \tag{3.24}$$

From this equation and $A > 0$, we find

$$W_{AP}P - W_{AB}B + t_A > 0 \Leftrightarrow B < \frac{W_{AP}P + t_A}{W_{AB}}, \tag{3.25}$$

which is the same as Eq. (3.7) in the main text. This condition defines a half-plane in the B-P space, delimited by a line with positive slope (the W_{IJ} are all positive). The nullclines and fixed points we derive below are only well-defined within this region (depicted as a green shaded area in Fig. 3.3A in the main text).

B nullcline in B-P space. To compute the B nullcline, we insert Eq. (3.24) in Eq. (3.3b) (in the steady-state):

$$B = f_B\left(W_{BP}P - W_{BB}B - \frac{k_A W_{BA}}{1 + k_A W_{AA}}(W_{AP}P - W_{AB}B + t_A)\right). \tag{3.26}$$

To solve for B and compute the non-zero portion of the B nullcline, we assume that the argument of f_B is larger than the threshold (see Eq. (3.5)), namely

$$W_{BP}P - W_{BB}B - W_{BA}A \geq -t_B. \tag{3.27}$$

Using Eq. (3.24), we obtain

$$W_{BP}P - W_{BB}B - \frac{k_A W_{BA}}{1 + k_A W_{AA}}(W_{AP}P - W_{AB}B + t_A) \geq -t_B. \tag{3.28}$$

Grouping the terms in Eq. (3.26), assuming that Eq. (3.28) holds, and solving for B yields:

$$\begin{aligned}\underbrace{\left(1 + k_B W_{BB} - \frac{k_B k_A W_{BA} W_{AB}}{1 + k_A W_{AA}}\right)}_{=:l} B = &\underbrace{\left(k_B W_{BP} - \frac{k_B k_A W_{BA} W_{AP}}{1 + k_A W_{AA}}\right)}_{=:\gamma} P \\ &+ \underbrace{\left(k_B \left(-\frac{k_A t_A W_{BA}}{1 + k_A W_{AA}} + t_B\right)\right)}_{=:m}\end{aligned} \tag{3.29}$$

so that, if $l \neq 0$, one branch of the B nullcline can be written as:

$$B = \frac{1}{l}(\gamma P + m). \tag{3.30}$$

If $l = 0$, Eq. (3.29) describes the vertical line $P = -\frac{m}{\gamma}$. The term γ can be interpreted as the difference in strength between the pathway $P \to B$ and the pathway $P \to \overset{\curvearrowleft}{A} \to B$ (including the recurrent connection W_{AA}). To comply with requirement 2 in Sec. 2.3.1, we need $\gamma > 0$. The term l can be thought of as the difference in strength between the recurrent connectivity in B and the pathway $B \to \overset{\curvearrowleft}{A} \to B$. The term m can be interpreted as the offset between a baseline input arriving at B through the pathway $A \to B$ and the threshold in B. At this point, we do not impose any constraints on the signs of l and m.

The portion of the B nullcline presented in Eq. (3.30) is well-defined when Eq. (3.28) holds. Grouping the terms and using the newly defined variables l, m and γ, Eq. (3.28) becomes

$$(1 - l)B + \gamma P + m \geq 0. \tag{3.31}$$

The equation defines a half-plane in the B-P space, delimited by a line with slope $s = -\frac{\gamma}{1-l}$, whose sign depends on the sign of the denominator. This also affects the direction of the inequality. Outside of the half-plane, the B nullcline is defined as $B = 0$. Thus, as described in the main text (Eq. (3.9)), the B nullcline can be written as

$$B = \begin{cases} \frac{1}{l}(\gamma P + m) & \text{if } W_{BP}P - W_{BB}B - W_{BA}A \geq -t_B \\ 0 & \text{else} \end{cases} \tag{3.32}$$

The line delimiting the half-space and Eq. (3.30) intersect at the point $P = -\frac{m}{\gamma}, B = 0$, where the constant and non-zero branches of the nullcline meet. Figure 3.3A in the main text presents an example where the B nullcline (blue line) is defined as Eq. (3.30) inside the half-plane defined by Eq. (3.31) (portion hatched with blue lines) and as $B = 0$ outside (to satisfy Eq. (3.5)).

P nullcline in B-P space. To compute the P nullcline, we insert Eq. (3.24) in Eq. (3.3a) (in the steady-state):

$$P = f_P\left(W_{PP}P - W_{PB}B - \frac{k_A W_{PA}}{1 + k_A W_{AA}}(W_{AP}P - W_{AB}B + t_A)\right). \tag{3.33}$$

To solve for B and compute the non-zero portion of the P nullcline, we have to assume that the argument to f_P is larger than the threshold (see Eq. (3.4)), namely

$$W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P. \tag{3.34}$$

Using Eq. (3.24), we obtain

$$W_{PP}P - W_{PB}B - \frac{k_A W_{PA}}{1 + k_A W_{AA}}(W_{AP}P - W_{AB}B + t_A) \geq -t_P. \tag{3.35}$$

Grouping the terms in Eq. (3.33), assuming that Eq. (3.23) holds, and solving for B yields:

$$\begin{aligned}\underbrace{\left(\frac{k_P k_A W_{PA} W_{AB}}{1 + k_A W_{AA}} - k_P W_{PB}\right)}_{=:c} B = &\underbrace{\left(1 - k_P W_{PP} + \frac{k_P k_A W_{PA} W_{AP}}{1 + k_A W_{AA}}\right)}_{=:g} P \\ &+ \underbrace{\left(k_P\left(\frac{k_A t_A W_{PA}}{1 + k_A W_{AA}} - t_P\right)\right)}_{=:h}\end{aligned} \tag{3.36}$$

and the P nullcline can be written as:

$$P = \frac{1}{g}(cB - h). \tag{3.37}$$

The term c can be interpreted as the difference in strength between the pathway $B \to \overset{\curvearrowleft}{A} \to P$ and the pathway $B \to P$. To comply with requirement 3 in Sec. 2.3.1, we need to impose $c > 0$. The term g can be thought of as the summed strength of the recurrent connectivity in P and the pathway $P \to \overset{\curvearrowleft}{A} \to P$. We assume that $1 - k_P W_{PP} > 0$, i.e., we assume that recurrent excitatory connections are small enough to avoid activity explosion, which implies $g > 0$. As a result, the non-zero portion of the P nullcline defined by Eq. (3.37) has a positive slope. The term h can be interpreted as the offset between a baseline input arriving at P through the pathway $A \to P$ and the threshold in P. We do not impose any constraint on its sign.

The portion of the P nullcline derived in Eq. (3.37) is well-defined when Eq. (3.35) holds. Grouping the terms and using the newly defined variables c, g and h, Eq. (3.35) results in

$$(1 - g)P + cB - h \geq 0. \tag{3.38}$$

This equation defines a half-plane in the B-P space; because $c > 0$, it can be rewritten as

$$B \geq \frac{(g-1)P + h}{c}. \tag{3.39}$$

The line delimiting the half-plane and Eq. (3.37) intersect at the point $P = 0, B = \frac{h}{c}$. Thus, for values in the half-plane, the P nullcline is defined by Eq. (3.37), and, outside of it, it must hold that $P = 0$ (to satisfy Eq. (3.4)). Thus, as described in the main text (Eq. (3.11)), the P nullcline can be defined as

$$P = \begin{cases} \frac{1}{g}(cB - h) & \text{if } W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P \\ 0 & \text{else.} \end{cases} \tag{3.40}$$

This is visualized in Fig. 3.3A of the main text, where the P nullcline in the B-P space (red line in B-P space) is non-zero inside the half-plane (portion hatched by red circles) and equals $P = 0$ outside of it.

Conditions for nullclines' intersections in the B-P space. Now that we have defined the P and B nullclines in the B-P space, we are interested in understanding when they have exactly two intersections in the first quadrant (corresponding to FP0 and FP2). We have already discussed that the following conditions need to be fulfilled:

- $\gamma > 0$, $c > 0$ to fulfill requirements 2 and 3, respectively (see Sec. 2.3.1)
- $1 - k_P W_{PP} > 0$, and thus $g > 0$, i.e. recurrent excitatory connections are small enough
- Eq. (3.25) is satisfied for all FPs

By definition FP0 should have $B = 0$, so it should belong to the constant portion of the B nullcline. Its P coordinate can be zero or positive. For FP2, both coordinates should be positive.

To understand under which conditions the nullclines intersect in exactly two points, we distinguish two cases: $l \geq 0$ and $l < 0$. In Fig. 3.14, we visualize the nullclines in the different cases. Note that $\gamma > 0$ in Eq. (3.30) and $c > 0$, $g > 0$ in Eq. (3.37).

Case $l \geq 0$

- If $l > 0$, the slopes of Eqs. (3.30), (3.37) are both positive. For the P nullcline to intersect the B nullcline in exactly two points, the slope of the P nullcline should be shallower than that of the B nullcline, yielding $gl - \gamma c < 0$. If $l = 0$, the branch of the B nullcline is a vertical line, and thus $gl - \gamma c < 0$ is always satisfied.
- For FP0 and FP2 to have non-negative P and B coordinates, one for each branch of the B nullcline, it should hold that $0 < -\frac{m}{\gamma}$. From $\gamma > 0$ it follows that $m < 0$.
- If $h < 0$, FP0 will have positive P coordinate ($P = -\frac{h}{g}$); if $h \geq 0$, then the P coordinate is $P = 0$.

Case $l < 0$

- If $l < 0$, the slope of Eq. (3.30) is negative, whereas the slope of Eq. (3.37) is positive (because $c > 0$, $g > 0$). Thus, the lines have opposite slopes and $gl - \gamma c < 0$ is always satisfied.

- To impose conditions for FP0 and FP2 having non-negative coordinates, we study the two cases $h < 0$, $h \geq 0$.
  - If $h < 0$, then the P nullcline intersects $B = 0$ at the point $P = -\frac{h}{g}$ (which is positive). To make sure that the B nullcline is defined at this point (so that the point is FP0), we need to impose $-\frac{h}{g} < -\frac{m}{\gamma}$. This implies that $m < 0$.
  - If $h \geq 0$, we need to impose that $0 < -\frac{m}{\gamma}$ (to guarantee that the B nullcline is defined in the first quadrant, which implies $m < 0$). Additionally, we need to make sure that the intersection of lines defined by Eqs. (3.30), (3.37) is positive (it yields FP2). This point has P coordinate $P = \frac{cm-lh}{gl-c\gamma}$, and, as $gl - c\gamma < 0$, we need to impose that $cm - lh < 0$.

3.6.3.2 A-P space

We now restrict ourselves to the 2d space A-P, assuming that $B > 0$ (i.e. we exclude the case in which $B = 0$). This allows us to solve Eq. (3.3b) for B and to use it to compute the P and A nullclines in this space. The nullclines should intersect in two points, corresponding to the SWR fixed point (FP1) and the middle FP (FP2). Both FPs have a non-zero B coordinate, whereas FP0, not visible in this space, has $B = 0$ by construction (and was already visualized in the B-P space discussed in Sec. 3.6.3.1).

Assuming that $B > 0$ means assuming that the argument of f_B (defined in Eq. (3.5)) is larger than the threshold. Thus, we assume that

$$W_{BP}P - W_{BB}B - W_{BA}A \geq -t_B. \tag{3.41}$$

If this is the case, we can solve Eq. (3.3b) for B (in the steady-state, with non-zero branch of Eq. (3.5)) and obtain

$$B = \frac{k_B}{1 + k_B W_{BB}}(W_{BP}P - W_{BA}A + t_B). \tag{3.42}$$

From this equation, and $B > 0$, we find

$$W_{BP}P - W_{BA}A + t_B > 0 \Leftrightarrow A < \frac{W_{BP}P + t_B}{W_{BA}}, \tag{3.43}$$

which is the same as Eq. (3.12) in the main text. This condition defines a half-plane in the A-P space, delimited by a line with positive slope (as the W_{IJ} are all positive). The nullclines and fixed points we are deriving below are only well-defined within this region (depicted as a blue shaded area in Fig. 3.3B).

A nullcline in A-P space. We follow an approach analogous to what we have explained in the B-P space to derive the A and P nullclines. Nevertheless, it is important to present the derivations in the A-P space, as the conditions on the strength of the pathways discussed in Sec. 2.3.1 impose a different scenario than the one we have discussed

in the B-P space. To compute the A nullcline, we insert Eq. (3.42) into Eq. (3.3c):

$$A = f_A\left(W_{AP}P - \frac{k_B W_{AB}}{1 + k_B W_{BB}}(W_{BP}P - W_{BA}A + t_B) - W_{AA}A\right). \tag{3.44}$$

To compute the non-zero portion of the A nullcline, we assume that

$$W_{AP}P - W_{AB}B - W_{AA}A \geq -t_A. \tag{3.45}$$

Using Eq. (3.42), we obtain

$$W_{AP}P - \frac{k_B W_{AB}}{1 + k_B W_{BB}}(W_{BP}P - W_{BA}A + t_B) - W_{AA}A \geq -t_A. \tag{3.46}$$

Grouping the terms in Eq. (3.44), assuming that Eq. (3.46) holds and solving for A yields

$$\begin{aligned}\underbrace{\left(1 + k_A W_{AA} - \frac{k_A k_B W_{AB} W_{BA}}{1 + k_B W_{BB}}\right)}_{=:l'} A = &\underbrace{\left(k_A W_{AP} - \frac{k_A k_B W_{AB} W_{BP}}{1 + k_B W_{BB}}\right)}_{=:\gamma'} P \\ &+ \underbrace{\left(k_A\left(-\frac{k_B t_B W_{AB}}{1 + k_B W_{BB}} + t_A\right)\right)}_{=:m'}.\end{aligned} \tag{3.47}$$

The term γ' can be interpreted as the difference in strength between the pathway $P \to A$ and the pathway $P \to \overset{\curvearrowleft}{B} \to A$. To comply with requirement 1 in Sec. 2.3.1, we need to impose $\gamma' < 0$. Note that the corresponding variable in the B-P space was $\gamma > 0$. The term l' can be thought of as the difference in strength between the recurrent connectivity in A and the pathway $A \to \overset{\curvearrowleft}{B} \to A$. Note that

$$l' = \frac{l(1 + k_A W_{AA})}{1 + k_B W_{BB}}, \tag{3.48}$$

and l' will therefore have the same sign as l. The term m' can be interpreted as the offset between a baseline input arriving to A through the pathway $B \to A$ and the threshold in A. Initially, we do not impose any constraints on the signs of l' and m'.

Assuming that $l' \neq 0$, the non-zero portion of the A nullcline can be rewritten as

$$A = \frac{1}{l'}(\gamma' P + m'), \tag{3.49}$$

and is well-defined when Eq. (3.46) holds. If $l' = 0$, the equation describes the vertical line $P = -\frac{m'}{\gamma'}$. Grouping the terms and using the newly defined variables l', γ', and m', Eq. (3.46) becomes

$$\gamma' P + (1 - l')A + m' \geq 0. \tag{3.50}$$

This equation defines a half-plane in the A-P space, delimited by a line with slope $s = -\frac{\gamma'}{1-l'}$, whose sign depends on the sign of the denominator. This also affects the direction of the inequality. Outside of this half-plane, the A nullcline is defined as $A = 0$.

As described in the main text (Eq. (3.14)), the A nullcline can be defined as

$$A = \begin{cases} \frac{1}{l'}(\gamma' P + m') & \text{if } W_{AP}P - W_{AB}B - W_{AA}A \geq -t_A \\ 0 & \text{else} \end{cases} \tag{3.51}$$

The line delimiting the half-space and Eq. (3.49) intersect at the point $P = -\frac{m'}{\gamma'}, A = 0$, where the constant and non-zero branches of the nullcline meet. Figure 3.3B presents an example where the A nullcline (green line) is defined as Eq. (3.49) inside the half-plane defined by Eq. (3.50) (portion hatched with green circles) and as $A = 0$ outside of it (to satisfy Eq. (3.6)).

P nullcline in A-P space. Finally, we turn to the derivation of the P nullcline in the A-P space. First, we insert Eq. (3.42) in Eq. (3.3a) (in the steady-state):

$$P = f_P\left(W_{PP}P - \frac{W_{PB}k_B}{1 + k_B W_{BB}}(W_{BP}P - W_{BA}A + t_B) - W_{PA}A\right). \tag{3.52}$$

We assume that the argument of f_P is larger than the threshold (see Eq. (3.4)), namely

$$W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P. \tag{3.53}$$

Using Eq. (3.42), we obtain

$$W_{PP}P - \frac{W_{PB}k_B}{1 + k_B W_{BB}}(W_{BP}P - W_{BA}A + t_B) - W_{PA}A \geq -t_P. \tag{3.54}$$

Grouping the terms in Eq. (3.52), assuming Eq. (3.54) holds, and solving for A yields:

$$\begin{aligned} \underbrace{\left(\frac{k_P k_B W_{PB} W_{BA}}{1 + k_B W_{BB}} - k_P W_{PA}\right)}_{=:c'} A = &\underbrace{\left(1 - k_P W_{PP} + \frac{k_P k_B W_{PB} W_{BP}}{1 + k_B W_{BB}}\right)}_{=:g'} P \\ &+ \underbrace{\left(k_P\left(\frac{k_B t_B W_{PB}}{1 + k_B W_{BB}} - t_P\right)\right)}_{=:h'} \end{aligned} \tag{3.55}$$

so that the non-zero portion of the P nullcline can be written as:

$$P = \frac{1}{g'}(c'A - h'). \tag{3.56}$$

The term c' can be interpreted as the difference in strength between the pathway $A \to \overset{\curvearrowleft}{B} \to P$ and the pathway $A \to P$. To comply with requirement 5 in Sec. 2.3.1, we need $c' < 0$. Note that the corresponding variable in the B-P space was $c > 0$. The term g' can be thought of as the summed strength of the recurrent connectivity in P and the pathway $P \to \overset{\curvearrowleft}{B} \to P$. As we have previously assumed (Sec. 3.6.3.1) that $1 - k_P W_{PP} > 0$, it follows that $g' > 0$. The term h' can be interpreted as the offset

between a baseline input arriving to P through the pathway $B \to P$ and the threshold in P. At this point, we do not impose any sign constraint on h'.

The portion of the P nullcline presented in Eq. (3.56) is well-defined when Eq. (3.54) holds. Grouping the terms and using the newly defined variables c', g' and h', Eq. (3.54) reads

$$(1 - g')P + c'A - h' \geq 0. \tag{3.57}$$

This equation defines a half-plane in the A-P space; as $c' < 0$, we can write

$$A \leq \frac{(g' - 1)P + h'}{c'}. \tag{3.58}$$

The line delimiting the half-plane and Eq. (3.56) intersect at the point $P = 0, A = \frac{h'}{c'}$. For values in the half-plane, the P nullcline is defined by Eq. (3.56), and $P = 0$ outside (to satisfy Eq. (3.4)). As described in the main text (Eq. (3.17)), the P nullcline can be defined as

$$P = \begin{cases} \frac{1}{g'}(c'A - h') & \text{if } W_{PP}P - W_{PB}B - W_{PA}A \geq -t_P \\ 0 & \text{else} \end{cases} \tag{3.59}$$

This can be visualized in Fig. 3.3B, where the P nullcline in the A-P space (red line in the right plots) is non-zero inside the half-plane (portion hatched by red circles) and would be $P = 0$ outside of it. We do not see the constant branch of the P nullcline because for the parameters combinations presented in Fig. 3.3, it lies outside of the domain of definition described by Eq. (3.43).

Conditions for nullcline intersections, A-P space. Now that the nullclines have been defined in the A-P space, we impose conditions for them having two intersections in the first quadrant (FP1 and FP2). We have already discussed that the following conditions need to be fulfilled:

- $\gamma' < 0$, $c' < 0$ to fulfill requirements 1 and 5, respectively (see Sec. 2.3.1)
- $g' > 0$, i.e., recurrent excitatory connections are small enough (always valid since we impose $1 - k_P W_{PP} > 0$ in the analysis of the B-P space)
- Eq. (3.43) is satisfied for all FPs

FP1 should by definition have $A = 0$, so it should belong to the constant portion of the A nullcline. Its P coordinate can be zero, or positive. For FP2, both coordinates should be positive.

To understand under which conditions the nullclines intersect in exactly two points, we need to distinguish the cases $l' \geq 0$, $l' < 0$. Note that due to Eq. (3.48), the two cases correspond to the two cases for l discussed in the B-P space. Moreover, note that $\gamma' < 0$ in Eq. (3.49) and $c' < 0$, $g' > 0$ in Eq. (3.56).

Case $l' \geq 0$

- If $l' > 0$, the slopes of Eqs. (3.49), (3.56) are both negative. For the P nullcline to intersect the A nullcline in exactly two points, the slope of the P nullcline should

be shallower that of the A nullcline, yielding $g'l' - \gamma' c' < 0$. If $l = 0$, the branch of the A nullcline is a vertical line and $g'l' - \gamma' c' < 0$ is always satisfied.

- For FP0 and FP2 to have non-negative P and A coordinates, one for each branch of the A nullcline, it should hold that $0 < -\frac{m'}{\gamma'} < -\frac{h'}{g'}$. This yields $m' > 0$. Additionally, we need to assure that the intersection of lines defined by Eqs. (3.49), (3.56) is positive (it yields FP2). This point has P coordinate $P = \frac{c'm' - l'h'}{g'l' - c'\gamma'}$, and, as $g'l' - c'\gamma' < 0$, we impose $c'm' - l'h' < 0$.
- The P coordinate of FP1 is positive ($P = -\frac{h'}{g'}$), i.e. $h' < 0$.

Case $l < 0$

- If $l' < 0$, the slope of Eq. (3.49) is positive, whereas Eq. (3.56) has a negative slope (since $c' < 0$, $g' > 0$). Thus, the lines have opposite slopes and $g'l' - c'\gamma' < 0$ is always fulfilled.
- The conditions for FP0 and FP2 having non-negative P and A coordinates are similar to those of the $l' \geq 0$ case, except that we can relax the constraints on $0 < -\frac{m'}{\gamma'}$, as far as FP2 has a positive P coordinate. Hence, we need to impose that $-\frac{m'}{\gamma'} < -\frac{h'}{g'}$ and $c'm' - l'h' < 0$.
- FP1 will have a positive P coordinate ($P = -\frac{h'}{g'}$), i.e. $h' < 0$.

3.6.3.3 Linear stability analysis

Now that the nullclines have been defined in both 2d subspaces B-P and A-P, we can use the newly defined variables c, g, h, etc. to express the FP equations as intersections of nullclines.

Non-SWR state (FP0):

$$\begin{aligned} P_0 &= \begin{cases} -\frac{h}{g} & \text{if } h < 0 \\ 0 & \text{else} \end{cases} \\ B_0 &= 0 \\ A_0 &= \frac{k_A W_{AP} P_0 + k_A t_A}{1 + k_A W_{AA}}, \end{aligned} \tag{3.60}$$

where P_0 is the intersection between the line $B = 0$ and Eq. (3.37); B_0 is zero by construction, and A_0 is computed by inserting P_0 and B_0 in Eq. (3.24). Because $P_0 \geq 0$, and all other terms are positive, it follows that A_0 is always positive.

SWR state (FP1):

$$\begin{aligned} P_1 &= -\frac{h'}{g'} \\ B_1 &= \frac{k_B W_{BP} P_1 + k_B t_B}{1 + k_B W_{BB}} \\ A_1 &= 0, \end{aligned} \tag{3.61}$$

where P_1 is the intersection between the line $A = 0$ and Eq. (3.56); A_1 is zero by construction, and B_1 is computed by inserting P_1 and A_1 in Eq. (3.42). Because $P_1 \geq 0$, and all other terms are positive, it follows that B_1 is always positive.

Middle FP (FP2):

$$\begin{aligned} P_2 &= \frac{cm - lh}{lg - c\gamma} = \frac{c'm' - l'h'}{l'g' - c'\gamma'} \\ B_2 &= \frac{gm - \gamma h}{gl - \gamma c} \\ A_2 &= \frac{g'm' - \gamma' h'}{g'l' - \gamma' c'}, \end{aligned} \tag{3.62}$$

where P_2 and B_2 are the intersection between Eq. (3.30) and Eq. (3.37); at the same time, P_2 can also be computed as the P coordinate of the intersection between the lines defined in Eq. (3.49) and Eq. (3.56). The A coordinate of this intersection point defines A_2.

Fixed points' stability. We now look at the stability of FPs and show that this depends on the newly defined variables g, g'. With linear stability analysis, we restrict ourselves to studying the signs of the eigenvalues of the Jacobian matrices of the system at FP0 (non-SWR state), FP1 (SWR state), and FP2 (middle FP). The Jacobian at the non-SWR state reads

$$J_{non-SWR} = \begin{pmatrix} \frac{-1+k_P W_{PP}}{\tau_P} & \frac{-k_P W_{PB}}{\tau_P} & \frac{-k_P W_{PA}}{\tau_P} \\ 0 & \frac{-1}{\tau_B} & 0 \\ \frac{k_A W_{AP}}{\tau_A} & \frac{-k_A W_{AB}}{\tau_A} & \frac{-1-k_A W_{AA}}{\tau_A} \end{pmatrix}, \tag{3.63}$$

where the second row has zeros entries because $f_B(x) = 0$ (see Eq. (3.5)). The characteristic equation $\det(J_{non-SWR} - \lambda I) = 0$ yields

$$\left(-\frac{1}{\tau_B} - \lambda\right) \det \begin{pmatrix} \frac{-1+k_P W_{PP}}{\tau_P} - \lambda & \frac{-k_P W_{PA}}{\tau_P} \\ \frac{k_A W_{AP}}{\tau_A} & \frac{-1-k_A W_{AA}}{\tau_A} - \lambda \end{pmatrix} = 0 \tag{3.64}$$

Thus, $\lambda_1 = -\frac{1}{\tau_B}$ is one negative eigenvalue. To determine the sign of the other eigenvalues we note that the determinant in Eq. (3.64) is $\det(J_{22} - \lambda I)$, where J_{22} is the 2×2 submatrix of $J_{non-SWR}$ obtained by deleting the central diagonal entry. Eigenvalues of this matrix are both negative if and only if $\mathrm{Tr}(J_{22}) < 0$ and $\det(J_{22}) > 0$. This results in the following conditions:

$$\begin{aligned} \mathrm{Tr}(J_{non-SWR}) < 0 \quad &\Leftrightarrow \quad \frac{1 + k_A W_{AA}}{\tau_A} + \frac{1 - k_P W_{PP}}{\tau_P} > 0 \\ \det(J_{non-SWR}) > 0 \quad &\Leftrightarrow \quad x := (1 - k_P W_{PP})(1 + k_A W_{AA}) + k_A k_P W_{AP} W_{PA} > 0 \end{aligned} \tag{3.65}$$

The condition on the trace is fulfilled as $1 - k_P W_{PP} > 0$ (see Sec. 3.6.3.1). Because $x = g(1 + k_A W_{AA})$ and $g > 0$, the condition on the determinant is also fulfilled in our scenario, so FP0 is stable.

Similarly, for the SWR state, the Jacobian yields

$$J_{SWR} = \begin{pmatrix} \frac{-1+k_P W_{PP}}{\tau_P} & \frac{-k_P W_{PB}}{\tau_P} & \frac{-k_P W_{PA}}{\tau_P} \\ \frac{-k_B W_{BP}}{\tau_B} & \frac{-1-k_B W_{BB}}{\tau_B} & \frac{-k_B W_{BA}}{\tau_B} \\ 0 & 0 & \frac{-1}{\tau_A} \end{pmatrix}, \tag{3.66}$$

where the third row has zeros entries because $f_A(x) = 0$ (see Eq. (3.6)). The stability conditions are:

$$\begin{aligned} \mathrm{Tr}(J_{SWR}) < 0 \quad &\Leftrightarrow \quad \frac{1 + k_B W_{BB}}{\tau_B} + \frac{1 - k_P W_{PP}}{\tau_P} > 0 \\ \det(J_{SWR}) > 0 \quad &\Leftrightarrow \quad x' := (1 - k_P W_{PP})(1 + k_B W_{BB}) + k_B k_P W_{BP} W_{PB} > 0. \end{aligned} \tag{3.67}$$

The condition on the trace is fulfilled as $1 - k_P W_{PP} > 0$. As $x' = g'(1 + k_A W_{AA})$, the condition on the determinant is also fulfilled in our scenario, so FP1 is a stable point.

We now consider the Jacobian matrix for the middle point (where both B and A coordinates are non-zero):

$$J_{middle} = \begin{pmatrix} \frac{-1+k_P W_{PP}}{\tau_P} & \frac{-k_P W_{PB}}{\tau_P} & \frac{-k_P W_{PA}}{\tau_P} \\ \frac{-k_B W_{BP}}{\tau_B} & \frac{-1-k_B W_{BB}}{\tau_B} & \frac{-k_B W_{BA}}{\tau_B} \\ \frac{k_A W_{AP}}{\tau_A} & \frac{-k_A W_{AB}}{\tau_A} & \frac{-1-k_A W_{AA}}{\tau_A} \end{pmatrix}. \tag{3.68}$$

We require that the FP2 is unstable, to represent a threshold between the two stable points. Every time the system is able to jump over this point, a transition (from non-SWR state to SWR state, or vice versa) will happen. We note that

$$\mathrm{Tr}(J_{middle}) = -\left(\frac{1 - k_P W_{PP}}{\tau_P} + \frac{1 + k_B W_{BB}}{\tau_B} + \frac{1 + k_A W_{AA}}{\tau_A} \right) < 0, \tag{3.69}$$

as $1 - k_P W_{PP} > 0$. Thus, the matrix has at least one negative eigenvalue. For the FP2 to be unstable, at least one eigenvalue should be positive. Two cases are possible: if the signs of the eigenvalues are $++-$, $\det(J_{middle}) < 0$; otherwise, if the signs are $+--$, $\det(J_{middle}) > 0$. Hence, we cannot use any argument on the determinant to constrain our search. Unfortunately, the explicit calculation of the eigenvalues is prohibitive, and we have to resort to numerical methods to extract their sign and check that at least one positive and one negative eigenvalues exist.

3.6.3.4 Summary of derived conditions

We summarize below the conditions to be enforced for the system to have three FPs with the right stability (they correspond to the conditions presented in Sec. 3.3.1.4). These conditions are a summary of the conditions for the nullclines to intersect in the B-P and A-P subspaces, and the stability conditions. Additionally, we impose that $P_0 < P_1$ to match experimental results.

1. $c > 0$, $\gamma > 0$, $c' < 0$, $\gamma' < 0$ (requirements on pathways' strength)
2. $1 - k_P W_{PP} > 0$ (small excitatory recurrent connections; it implies $g > 0$, $g' > 0$)
3. if $l > 0$, impose $gl - \gamma c < 0$ (slope condition in B-P space, always holds for $l \leq 0$)
4. if $l \geq 0$, it should hold $0 < -\frac{m}{\gamma}$. If $l < 0$ and $h < 0$, we need to impose $-\frac{h}{g} < -\frac{m}{\gamma}$, whereas if $l < 0$ and $h \geq 0$, it should hold that $0 < -\frac{m}{\gamma}$ and that $cm - lh < 0$ (ordered intersections at $B = 0$, non-negative FPs)
5. $B_2 < \frac{W_{AP}P_2 + t_A}{W_{AB}}$ (for Eq. (3.25) to be fulfilled at FP2)
6. if $l' > 0$, impose $g'l' - \gamma' c' < 0$ (slope condition in A-P space, always holds if $l' \leq 0$)
7. $-\frac{m'}{\gamma'} < -\frac{h'}{g'}$ and $c'm' - l'h' < 0$ (for ordered intersections at $A = 0$ and non-negative FPs; it implies $h' < 0$). If $l' \geq 0$, it additionally holds that $m' > 0$
8. $A_2 < \frac{W_{BP}P_2 + t_B}{W_{BA}}$ (for Eq. (3.43) being fulfilled at FP2)
9. $hg' - h'g > 0$ (to have $P0 < P1$)
10. J_{middle} has at least one positive eigenvalue (numerical)

Conditions 1 (on c and γ) and 2-4 are derived in Sec. 3.6.3.1. In that section, we also mentioned that Eq. (3.25) should be fulfilled for all fixed points. It is easy to show that the condition is fulfilled for FP0. In fact:

$$W_{AP}P_0 - W_{AB}B_0 + t_A > 0 \Leftrightarrow W_{AP}P_0 + t_A > 0, \tag{3.70}$$

which is fulfilled as P_0 is non-negative. For the other fixed point in the B-P space (FP2), we explicitly need to enforce the condition, which we do in condition 5. Conditions 1 (on c', γ'), and 6-7 are derived in Sec. 3.6.3.2. As mentioned in that section, we need to check that Eq. (3.43) is fulfilled by the FPs. For FP1, it holds

$$W_{BP}P_1 - W_{BA}A_1 + t_B > 0 \Leftrightarrow W_{BP}P_1 + t_B > 0, \tag{3.71}$$

because P_1 is positive. For FP2, we explicitly impose condition 8. Condition 9 has been imposed to fulfill experimental results on pyramidal cells increasing their firing during SWRs (see Sec. 1.2.3), and condition 10 is derived in Sec. 3.6.3.3.

Fig. 3.14 shows the nullclines in the B-P and A-P spaces, for different combinations of l and h. The figure shows that large portions of the nullclines (and their intersections, the FPs) are included in the region where conditions 5 and 8 hold (shaded areas), and are thus well-defined.

3.6.4 Definition of new spiking models

Neuronal parameters. I create a new spiking network with conductance-based LIF units. The intrinsic neuronal parameters are taken from the spiking model presented in Chapter 2 and are listed in Table 2.1.

Connections. The connections strengths W_{IJ} are parameters of the mean-field network, which I use to define connection parameters in the spiking model. The link between

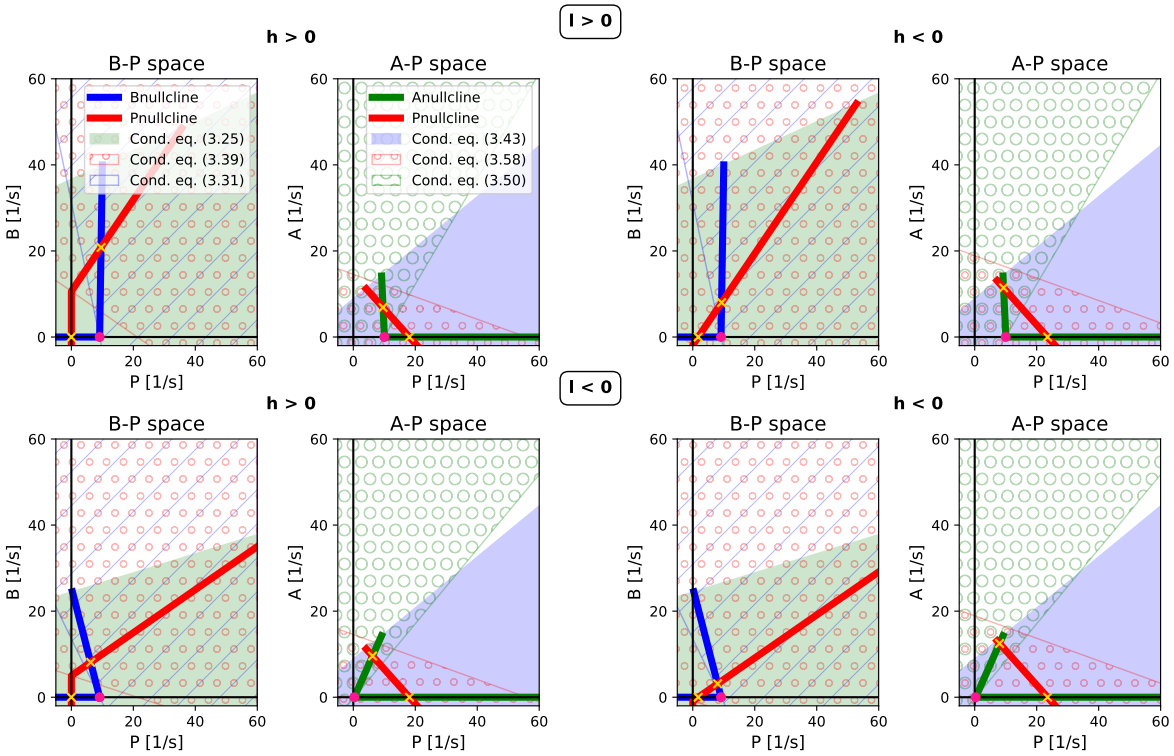

FIGURE 3.14: **Nullclines in B-P and A-P spaces across different conditions.** Each set of two plots shows the nullclines in the 2d subspaces for different values of W_{AB} and t_P, resulting in sign changes of l, h. Left: P (red) and B (blue) nullclines in the B-P space. Green shaded area represents the portion where Eq. (3.25) holds; nullclines are only defined inside this area. The area hatched with red circles represents the portion where Eq. (3.39) holds, i.e., where the P nullcline is non-zero, and the area hatched with blue lines defines the portion where Eq. (3.31) holds, i.e., the B nullcline is non-zero. Outside these regions, nullclines are constant. Gold crosses are the FPs of the system (FP0 with $P_0 = \max(0, -h/g)$, $B_0 = 0$, and FP2). The pink dot represents the point $-m/\gamma$. Right: P (red) and A (green) nullclines in the A-P space. Blue shaded area represents the portion where Eq. (3.43) holds; nullclines are only defined inside this area. The area hatched with red circles represents the portion where Eq. (3.58) holds, i.e., where the P nullcline is non-zero, and the area hatched with green circles defines the portion where Eq. (3.50) holds, i.e., the A nullcline is non-zero. Outside these regions, nullclines are constant. Gold crosses are the FPs of the system (FP1 with $P_1 = -h'/g'$, $A_1 = 0$, and FP2), and the pink dot represents the point $-m'/\gamma'$. Parameters are varied in the four conditions as follows: $W_{AB} = 3$, $t_P = 115$ (upper left); $W_{AB} = 3$, $t_P = 135$ (upper right); $W_{AB} = 4.5$, $t_P = 135$ (lower right); $W_{AB} = 4.5$, $t_P = 115$ (lower left). Other network parameters are: $W_{PP} = 2$, $W_{BP} = 8$, $W_{PA} = 10$, $W_{PB} = 2$, $W_{BB} = 3$, $W_{BA} = 13.5$, $W_{AP} = 1$, $W_{AA} = 5$, $k_P = 0.4$, $k_B = 0.5$, $k_A = 0.3$, $t_B = 120$, $t_A = 110$, $\tau_P = 3$ ms, $\tau_B = 2$ ms, $\tau_A = 6$ ms.

these sets of variables has been explained in detail in Sec. 3.6.1. I rewrite here for completeness Eq. (3.22), focusing, on the connection $P \to B$:

$$W_{BP} = N_P p^{BP} g_{ij}^{BP} \tau_{syn}^{P} (E_{rev}^{P} - V_B), \tag{3.72}$$

where I assume that g_{ij}^{BP} are equal for all synapses (and drop the i, j dependence in what follows). Our aim is to extract the value of the product $p^{BP} g^{BP}$, given that the number N_P of P cells, the time constant τ_{syn}^{P} of AMPA synapses, and the reversal potential for excitation E_{rev}^{P} are fixed neuronal properties. In the next paragraph, I discuss how to deal with the term V_B, which describes the mean membrane potential value of the postsynaptic population. Assuming that we derive a value for $p^{BP} g^{BP}$, a choice still has to be made to assign a value to each factor. I assume that $p^{IJ} = 0.6, \forall I, J$ pairs and discuss how this choice impacts the results.

Optimization procedure on $\mathbf{V_I}$. To estimate the connectivity parameters p^{BP} and g^{BP} from Eq. (3.72), we need to choose a value for V_B. In a network with a single stationary state, V_B represents the mean membrane potential of population B in the steady-state. Its value depends on the properties of the neurons in the population and on the network-dependent input in a given state. Because the interaction between these factors is highly complex, the value of V_B cannot be determined *a priori* without simulating the spiking network. An additional complication in our scenario is that, as we are dealing with a bistable system, two stationary states coexist in the network, likely with two different mean membrane potentials for any given population. Figure 3.12 shows that this is indeed the case for the spiking model introduced in Chapter 2. To overcome the issue of how to define the mean membrane potential values V_I across states, I run an optimization procedure analogous to the one described in Sec. 3.6.1. I match the population rates in the mean-field model with the population firing rates of spiking models with fixed neuronal parameters and connectivity structure ($p^{IJ} = 0.6, \forall I, J$), and choose the combination of V_P, V_B, and V_A minimizing the distance between the firing rate vectors of spiking and mean-field models. More in detail, I explore the range $[-60, -50]$ mV in steps of 1 mV and simulate the spiking network with clamped synaptic efficacy $e_{IJ} = 0.5$. During each simulation, the network is initialized to start from a non-SWR state. Then, a depolarizing current is injected to P cells for the system to jump to the SWR state (if it exists), and population firing rates are recorded in each state. Figure 3.15 gives an overview of the population rates in the spiking model. The optimal values of V_P, V_B, and V_A are those which minimize the Euclidean norm between the target firing rates in the mean-field model and the population rates of the simulated spiking models. For the network presented in Fig. 3.8, optimal values are $V_P = -55$ mV, $V_B = -54$ mV, $V_A = -58$ mV. As a last step, g_{ij}^{IJ} can be defined using Eq. (3.22) because they are the only unknowns in the equations. Their values for the spiking network presented in Fig. 3.8 are listed in Table 3.2.

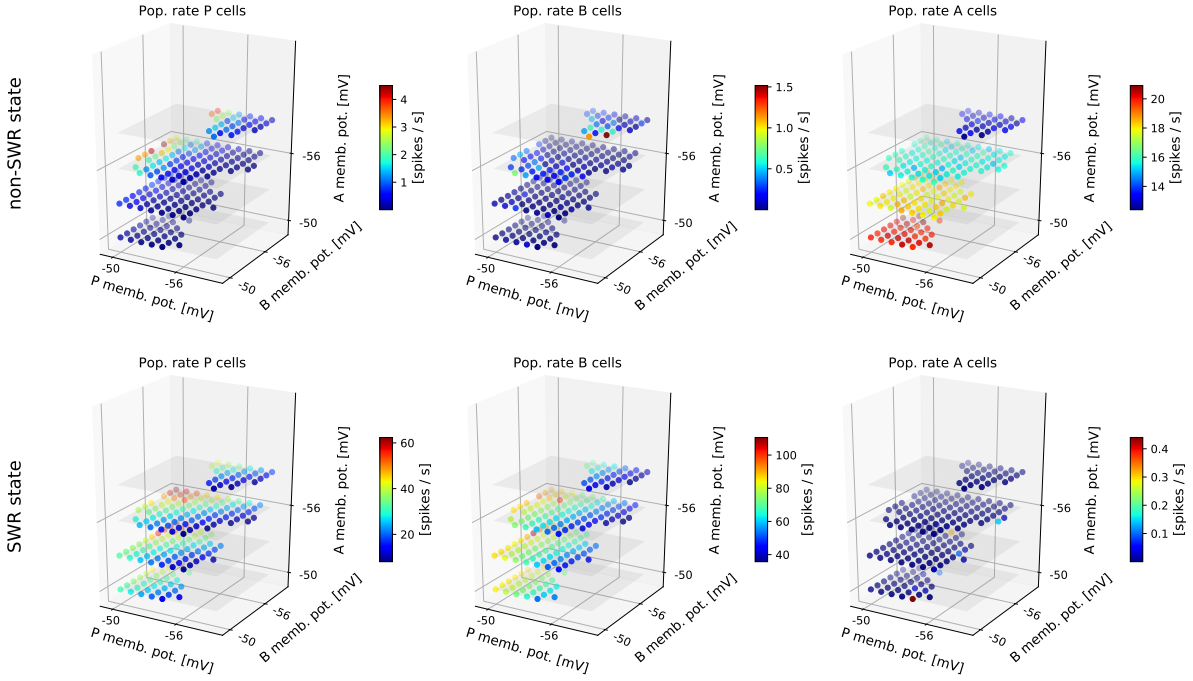

FIGURE 3.15: **Overview of population rates in spiking models with 60% connectivity and different values of $\mathbf{V_P}, \mathbf{V_B}, \mathbf{V_A}$.** Each panel shows the population firing rate value of a population in a given state (non-SWR, upper panels; SWR, lower panels) as a function of V_P, V_B, and V_A values (x-, y- and z-axis, respectively), in a network where the connection probability is $p = 0.6$ for all connections; neuronal parameters are default (see Table 2.1), and other parameters are derived from a mean-field model (displayed in Fig. 3.6, see Tables 3.2, 3.3). The choice $V_P = -55$ mV, $V_B = -54$ mV, and $V_A = -58$ mV yield the closest result (in terms of population firing rates) between the spiking and mean-field network. These values are used for the simulation of the spiking model in Fig. 3.8.

Spiking simulations. The simulations of the spiking model presented in Figs. 3.8, 3.9, and 3.10 are analogous to those described in Sec. 2.8. In Fig. 3.8, 60% of the cells are stimulated with current values uniformly distributed from 0 to a maximal value ($I_P = 300$ pA, $I_B = 500$ pA, and $I_A = \pm 800$ pA) for $T = 10$ ms. In Fig. 3.9, a negative current is injected to A cells at time 1 second to let the system jump to the SWR state. At time 4 seconds, positive current injection to A cells brings the system back to the non-SWR state. Finally, to generate Fig. 3.10, 50 cells per population are selected at random, and their response (one at a time) to current injection $[-100, 200]$ pA, with steps of 20 pA, is monitored (for 10 seconds). f-I curves are created as described in Sec. 2.8.5.

3.6.5 Spiking network parameters

The neuronal parameters for the spiking model shown in this chapter are the same as the ones used in Chapter 2 (see Table 2.1). Additionally, I list the connectivity values in the tables below.

Connection	Connection probability	Conductance update (nS)	Synaptic delay (ms)
P $\rightarrow$ P	0.6	$g^{PP} = 0.004$	$\tau_{PP} = 1$
P $\rightarrow$ A	0.6	$g^{AP} = 0.002$	$\tau_{AP} = 1$
A $\rightarrow$ A	0.6	$g^{AA} = 3.472$	$\tau_{AA} = 1$
A $\rightarrow$ P	0.6	$g^{PA} = 5.555$	$\tau_{PA} = 1$
P $\rightarrow$ B	0.6	$g^{BP} = 0.015$	$\tau_{BP} = 1$
B $\rightarrow$ B	0.6	$g^{BB} = 1.543$	$\tau_{BB} = 1$
B $\rightarrow$ P	0.6	$g^{PB} = 1.097$	$\tau_{PB} = 1$
A $\rightarrow$ B	0.6	$g^{BA} = 7.031$	$\tau_{BA} = 1$
B $\rightarrow$ A	0.6	$g^{AB} = 2.058$	$\tau_{AB} = 1$

TABLE 3.2: Values of the synaptic connection parameters for the spiking network shown in Figs. 3.8, 3.9, and 3.10.

Parameter	Value	Definition
η_D	0.4	Depression rate in connection $B \rightarrow A$
τ_D	300 ms	Synaptic depression time constant of connection $B \rightarrow A$

TABLE 3.3: Parameters of the synaptic depression mechanism of the spiking network shown in Figs. 3.8, 3.9 and 3.10.

4 | Identification of anti-SWR cells

The analysis of experimental data presented in this chapter is the result of a collaboration with the lab of Dietmar Schmitz, Charité Universitätsmedizin Berlin. The recordings were performed by Nikolaus Maier, Claire Cooper, and Aarti Swaminathan. They also contributed to the writing of Sec. 4.2 and 4.5.1.

4.1 Introduction

The models presented in Chapters 2 and 3 predict the existence of a class of inhibitory cells that are anti-modulated with respect to SWRs. As the cells are not yet identified, I termed them anti-SWR cells and assumed that they share the same neuronal properties (spiking threshold, resting potential, etc.) of the other two populations. Despite the generality of the model I am proposing and its independence from specific parameter values, the identification of anti-SWR cells would strongly support the model hypothesis. Furthermore, to test the causal link between anti-SWR cells and SWR generation, the selective activation or inactivation of this population via a targeted approach (e.g. optogenetics), would be necessary. In this chapter, I address these challenging questions and show the first evidence for the existence of anti-SWR cells in the hippocampal CA3 area.

In Section 1.2.3 I have reviewed the SWR-associated spiking patterns of the main types of identified interneuron classes and concluded that most show increased activity during SWRs. However, other classes of cells are not modulated by SWRs, and some cells (e.g. axo-axonic cells) seem to exhibit inconsistent behavior during SWRs (the variability possibly depends on the experimental conditions). Interestingly, Csicsvari et al. [1999b] identified, *in vivo*, CA1 interneurons that decreased their firing during SWR events. Figure 4.1A shows their firing distribution, temporally aligned with the ripple peak. These neurons were recorded in the alveus or *stratum oriens*, but no further details were provided about their identity. More recently, Fuentealba et al. [2008] reported the existence of an enkephalin-expressing (ENK) GABAergic cell in CA1, *in vivo*, which seemed to be anti-modulated with SWRs. Figure 4.1B shows its firing pattern with respect to ripples. More experiments are needed to confirm whether this is a general property of ENK-expressing cells. Finally, Le Van Quyen et al. [2008] showed that a subset of putative interneurons recorded in the human hippocampal formation stopped

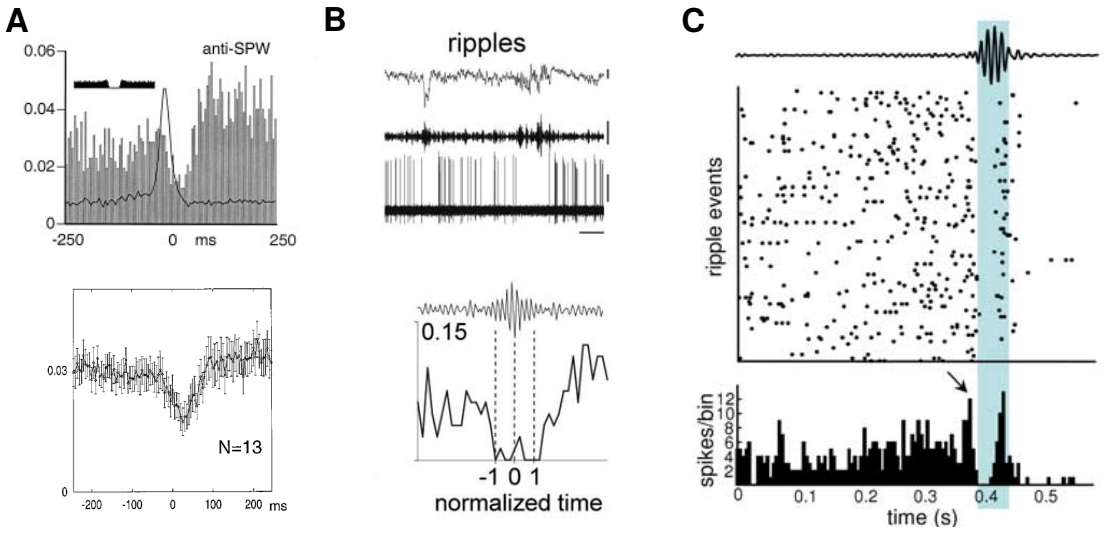

FIGURE 4.1: **Putative anti-SWR cells recorded in rodents and humans. A)** 'Anti-SPW' cells identified in rat CA1 (alveus and *stratum oriens*, $N = 13$). Top: discharge probability of a single anti-SPW cell aligned with ripple peak (time 0, gray histograms). The continuous line depicts the discharge of simultaneously recorded pyramidal cells. The inset shows cell's autocorrelogram (x-axis is 100 ms). Bottom: average discharge probability of multiple anti-SPW cells, aligned to ripple peak (time 0). The figure is reproduced from Csicsvari et al. [1999b] with permission from the Society for Neuroscience. **B)** ENK-expressing cell identified in rat CA1 (soma in *stratum radiatum*, axons projecting up to *stratum oriens*). Top: snapshot of recorded activity shows LFP (top), 90-140 Hz band-pass filtered signal to isolate ripple events (middle), and unit activity (bottom). Scale bars are: top, 0.5 mV; middle, 0.1 mV; bottom, 0.5 mV; horizontal, 500 ms. Bottom: firing probability aligned to the ripple peak (time 0). The figure is reproduced from Fuentealba et al. [2008] with permission from the Society for Neuroscience. **C)** Interneurons recorded in the human hippocampal formation that stop firing during SWRs ($N = 4$). Top: 80-200 Hz band-pass filtered LFP to reveal the ripple component; middle: raster plots of recorded cells across SWR sweeps; bottom: ripple-triggered spike histograms (bin size, 5 ms). Note the increased firing before and after the silent period. The figure is reproduced from Le Van Quyen et al. [2008] with permission from the Society for Neuroscience.

firing during the initial phase of a SWR event (Fig. 4.1C). Despite the evidence provided here, it remains to be determined whether interneurons with anti-modulated discharge properties also exist in the CA3 region.

4.2 Identification of anti-SWR cells in CA3, *in vitro*

We tested whether anti-SWR cells could be found in the CA3 region, which is known to be the site where SWRs are first generated ([Buzsáki, 1986; Maier et al., 2003; Ellender et al., 2010; Sullivan et al., 2011; Davoudi and Foster, 2019] and Sec. 1.1.3.2). Concretely, this required searching for interneurons exhibiting the predicted firing behavior, i.e., firing tonically outside of SWRs, while being suppressed during SWRs. This was achieved by means of an *in vitro* hippocampal slice preparation in which SWRs are observed spontaneously [Maier et al., 2003]. Briefly, extracellular spiking of putative hippocampal interneurons, identified primarily by their non-pyramidal soma shape, were targeted in

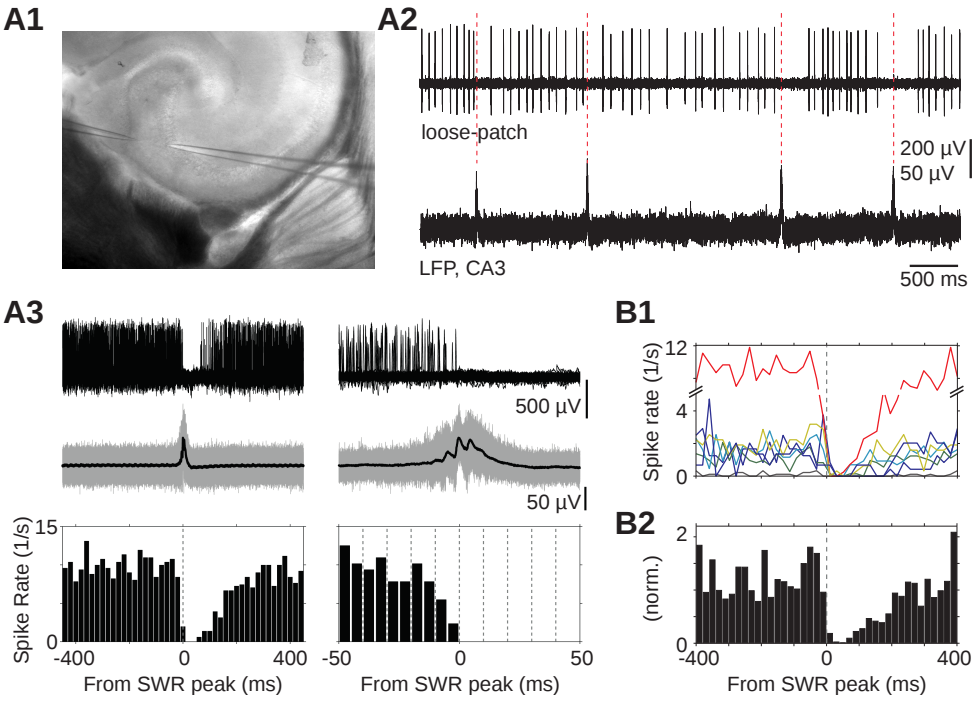

FIGURE 4.2: **Anti-SWR cells recorded in CA3. A1)** Infrared differential interference contrast (IR-DIC) video microscopy image of a slice with a patch and a LFP electrode. **A2)** SWRs in CA3 (bottom) in parallel with spiking of a recorded neuron (top). **A3)** Left, overlay of sweeps of 'loose-patch' recordings (top), simultaneously acquired SWRs (middle, 256 events, gray, and average, black) and respective spike time histograms (bottom, bin width 20 ms); same experiment as in A2. Right, same as left but at higher temporal resolution (bin width: 5 ms). **B1)** Overlay of spike rates (1/s) of anti-SWR cells ($N = 7$) with respect to SWR ripple peak. Each line represents the spike rate of one anti-SWR cell. **B2)** Normalized spike time histogram for all pooled anti-SWR cells (bin width: 20 ms). Normalization was performed using the median spike rate from 400 ms until 200 ms before the peak of the SWR.

the 'loose-patch' configuration (see Sec. 4.5.1). If the desired spiking behavior was observed, the candidate interneuron was then patched in the whole cell configuration and further characterized electrophysiologically. In total, seven putative anti-SWR interneurons were found. Figure 4.2A3 shows the firing histograms for one representative cell at different time resolutions, and Fig. 4.2B shows a summary of the activity of all recorded cells. An interesting observation is that anti-SWR cells are completely silent for up to 50 ms following the ripple peak, in agreement with what I have shown in the spiking model (see Fig. 2.2). Additionally, the activity in Fig. 4.2B2 shows a peak preceding the silent period, which is in line with the results of the spiking network when a SWR is initiated by activating P cells (because the direct connection $P \to A$ is faster than the disinhibitory pathway $B \to A \to P$, see Fig. 2.2).

Overall, these experimental results support the model hypothesis and should be considered a promising starting point for the characterization of anti-modulated cells in the CA3 region.

4.3 Does anti-SWR cell modulation affect SWRs?

The results of the previous section suggest that anti-SWR cells form a subset of CA3 interneurons. Therefore, they could contribute to the generation of SWRs as hypothesized by the model presented in Chapter 2. However, the SWR-associated modulation does not per se demonstrate that anti-SWR cells are causally implicated in SWR generation. At this point, we cannot exclude the hypothesis that they are modulated by SWRs without actively taking part in their generating mechanism. To better shed light on their contribution, a selective manipulation of the activity of anti-SWR cells is needed. If, for example, anti-SWR cells could be identified using genetic markers, optogenetic manipulation of their activity could be a useful tool to test the causality in SWR generation. Unfortunately, finding genetic markers specific to a given novel cell type is difficult. Thus, even though we hope that genetic markers for anti-SWR cells will become available in the future, we present here a simpler approach that is based on single cell manipulation. Under the assumption that anti-SWR cells represent a small fraction of hippocampal neurons, we test whether modifying the activity of a single cell might suffice to observe network-wide effects on, e.g., the SWR incidence.

4.3.1 Single cell manipulation

4.3.1.1 In the spiking model

To better understand the effect of single cell manipulation on the SWR behavior, I start by testing the effect of modifying the firing of a single anti-SWR cell (A in the model) in the spiking network. I reasoned that an effect could possibly be observed because only 50 cells are modeled in the network (see Sec. 2.8.2), and thus each cell strongly contributes to the generation of IPSPs at its postsynaptic targets.

To test the hypothesis, I perform two sets of analyses. First, I select one anti-SWR cell and inject constant hyperpolarizing current ($I = -500$ pA) for the whole duration of the simulation, during which I monitor the spontaneous sharp-wave-like LFP activity (see Sec. 2.8.4). The hyperpolarizing current effectively takes the cell out of the network, because the cell never fires over the course of the simulation. I repeat the procedure for each anti-SWR cell in the network and compare the SWR incidence in these simulations with the control condition (no cell manipulation, more details are provided in Sec. 4.5.2). Figure 4.3 shows that the SWR incidence tends to increase when a single anti-SWR cell is hyperpolarized (compare black and red histograms, samples not belonging to the same distribution, Wilcoxon rank-sum test, $p < 0.001$). This result is in line with the idea that less inhibition is present in the network, and it is thus easier for small fluctuations arriving to B cells to initiate a SWR event.

In the second set of experiments, I depolarize one single anti-SWR cell at a time. The choice of the constantly injected current $I = 150$ pA was based on the anti-SWR cells' f-I curves (Fig. 2.11) to make the cell more excitable (so that it fires more spikes than in the control conditions), but without being constantly active. The yellow histogram in

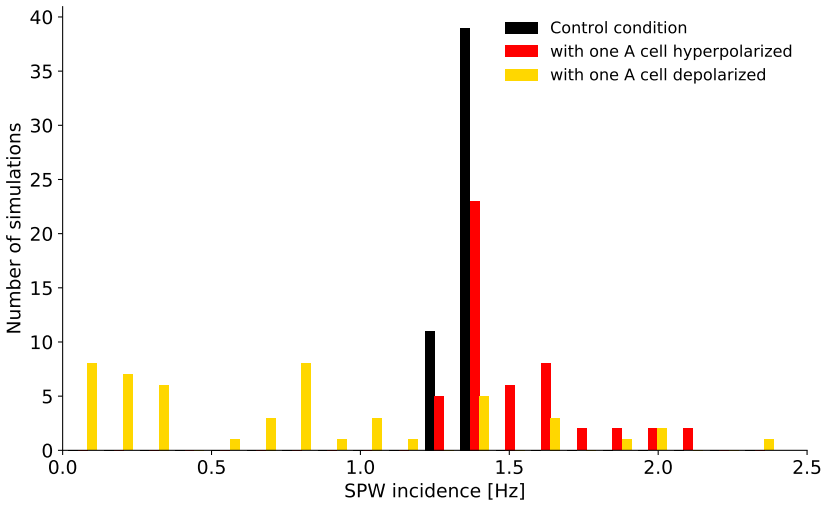

FIGURE 4.3: **Effect of single cell manipulation on the SWR incidence in the spiking network.** Histograms from 50 simulations of the spontaneous spiking network are shown. Colors refer to the control condition (black), the scenario where a single anti-SWR cell is hyperpolarized (red) or depolarized (yellow). The SWR incidence (x-axis) increases when anti-SWR cells are hyperpolarized. Anti-SWR cells depolarization can induce a decrease, an increase, or no changes of the SWR incidence.

Fig. 4.3 shows that the effect on SWR incidence is double-sided. In the majority of cases, the resulting SWR incidence is lower than in the control condition, but for some cells, the depolarization results in an increase of SWR incidence (compare black and gold histograms, samples not belonging to the same distribution, Wilcoxon rank-sum test, $p < 0.001$). This result suggests that the simple hypothesis of depolarizing anti-SWR cells resulting in more inhibition in the network, and thus less SWRs, needs to be revised.

To better understand the results of the simulations with depolarized anti-SWR cells, I reasoned that the net contribution of a cell to the network depends on the number of postsynaptic targets in the P, B, and A populations. My hypothesis was that if an anti-SWR cell contacts many other A cells, the effect of its depolarization in the network can be excitatory (as it inhibits many A cells), and hence, this depolarization can result in an increase in SWR incidence. In Figure 4.4, I show that the hypothesis is supported by a strong correlation between the number of postsynaptic targets onto A cells and the change in SWR incidence (right plot). Note that the number of targets is centered around the expected number of connections (size of postsynaptic population $\times$ connection probability, see Table 2.2), and that there is little correlation between SWR incidence and other types of postsynaptic targets.

As a control, I also plot the relation between SWR incidence and number of postsynaptic targets for the case in which an anti-SWR cell is hyperpolarized. Figure 4.5 shows that in this case, there is a less clear relation between the number of postsynaptic targets and the SWR incidence. However, even if the correlations are much lower in this case, this example confirms my hypothesis: if an A cell has few contacts to other A cells, but

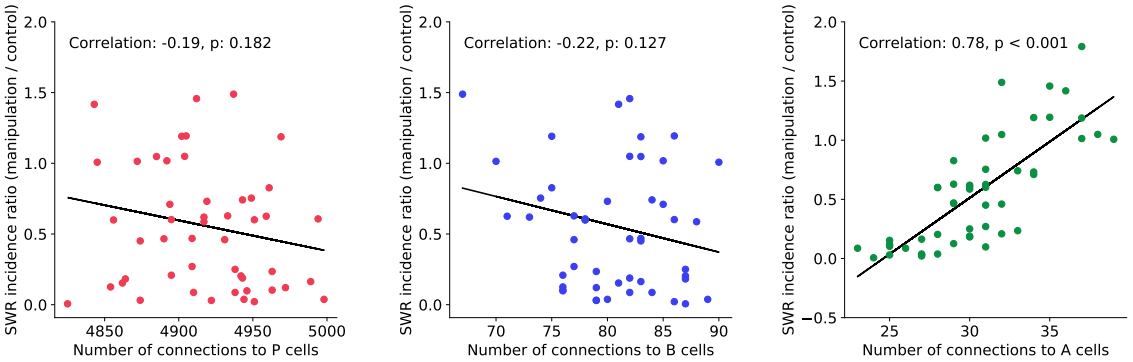

FIGURE 4.4: **The effect of anti-SWR cells depolarization correlates with the number of postsynaptic targets to the A population.** From left to right: the x-axis describes the number of postsynaptic targets in the P, B, and A population of each anti-SWR cell. The y-axis shows the ratio of SWR incidence in the simulations where a single anti-SWR cell was depolarized, with respect to the average incidence in the control condition. Each dot represents a simulation (one anti-SWR cell depolarized), black lines are best fit linear regression lines. Note the large positive correlation between SWR incidence and number of postsynaptic targets in the A population (right plot).

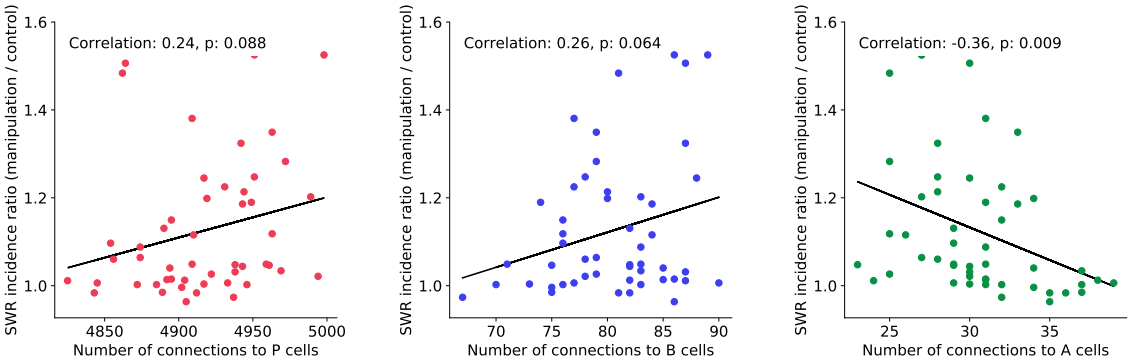

FIGURE 4.5: **The effect of anti-SWR cell hyperpolarization depends on multiple factors.** From left to right: the x-axis describes the number of postsynaptic targets in the P, B, and A population of each anti-SWR cell. The y-axis shows the ratio of SWR incidence in the simulations where a single anti-SWR cell was hyperpolarized, with respect to the average incidence in the control condition. Each dot represents a simulation (one anti-SWR cell hyperpolarized), black lines are best fit linear regression lines. In this scenario, all postsynaptic targets contribute to the amount of change of the SWR incidence.

many to P and B cells, its hyperpolarization increases the SWR incidence (compared to the control condition).

From this analysis, I predict that, if the number of anti-SWR cells is not too large in the biological network, the effect of single cell manipulation on the SWR incidence can be detected. The simulation results on single cell depolarization predict, however, that the effect of A cell activation might depend on the embedding of these cells in the network.

4.3.1.2 In the *in vitro* experiment

Now that we have analyzed *in silico* how single cell manipulation contributes to modifying SWR incidence, we can test if the results can be reproduced in the biological *in vitro* model. To this end, we use voltage clamp as a paradigm to hold the membrane potential of a single anti-SWR cell at a subthreshold value, and record SWR events using a LFP electrode. The voltage clamp condition should be comparable with the hyperpolarization condition in Sec. 4.3.1.1. The experimental setup is the same described in Sec. 4.5.1, and additional details on the data are provided in Sec. 4.5.3.

Figure 4.6 presents the preliminary results from five independent experiments. The SWR incidence is displayed as a function of the recording condition. From the results of Sec. 4.3.1.1, we predict that the SWR incidence increases when spiking in a single anti-SWR cell is suppressed, which we realize here by switching from current-clamp (control condition) to voltage-clamp mode. As we can see, only a subset of the recordings (2/5) seems to support this hypothesis. Note, however, that the range of SWR incidence change is relatively small. This can be explained by the fact that the activity of only a single anti-SWR cell is manipulated at a time. In addition, longer recordings and a more robust statistical analysis are required to differentiate any significant trend from measurement noise.

There are several reasons which could account for the discrepancy between the simulation and experimental results. First of all, the variability in recording conditions and the small sample complicate the interpretation of the results. Moreover, the technique needs to be improved to make sure that the paradigm minimizes time-dependent instabilities in the recording conditions (the interleaved condition used for cell 250319 seems a promising improvement, see Sec. 4.5.3). To have a fair comparison with the simulations, the recording time should be increased as much as possible. Inhomogeneities in the anti-SWR population not captured by the spiking model could also contribute to this discrepancy, and more information about the cell number, identity, and connectivity could be incorporated in the spiking model to test their influence on the analysis outcome.

4.4 Discussion

In this chapter, I have provided the first evidence for the existence of interneurons in the CA3 area that are anti-modulated by SWRs. It remains to be better understood whether these cells form a homogenous class, in terms of morphological properties, electrophysiological characteristics, or immunoreactivity to certain molecular markers. As more data will become available to better define the identity of these cells, it would be interesting to test how this group of cells relates to previously reported interneuron types, which have been shown to reduce their firing during SWRs (see Sec. 4.1 and Sec. 1.2.3). Note, in particular, that the model presented in Chapter 2 relies on the tonic firing of A cells in non-SWR periods to prevent P cells from spiking, a property that not all cells silent

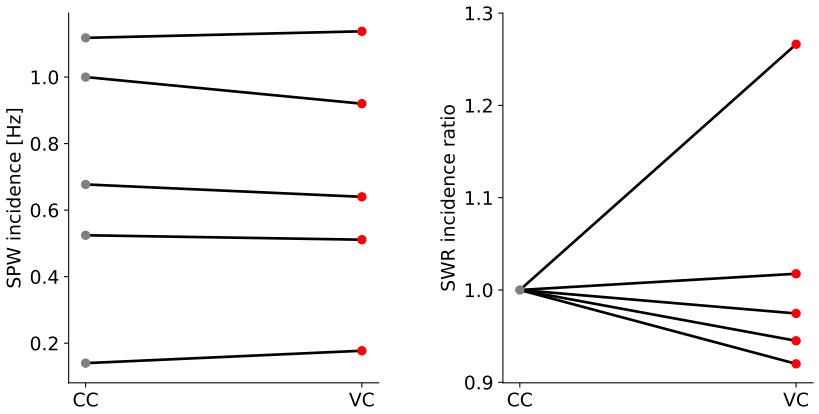

FIGURE 4.6: **Experimental results on single anti-SWR cell manipulation.** Results from 5 independent experiments are shown. Left: SWR incidence as a function of the recording conditions (CC: current clamp mode, control condition, gray dots; VC: voltage clamp mode, red dots, see Sec. 4.5.3). Note the large slice-to-slice variability in SWR incidence and the small changes induced by the VC condition. Right: Same recordings as left, normalized to show the change in the VC condition (compared to CC). Note that only two recordings show the predicted increase in SWR incidence when one anti-SWR cell is taken out of the network (VC condition, red dots).

during SWRs seem to fulfill (see Fig. 4.1 and Sec. 1.2.3). In this regard, it remains to be clarified in future experiments whether the tonic firing of anti-SWR cells is driven by intrinsic or network mechanisms.

Another interesting line of research deals with understanding how anti-SWR cells are embedded in the network. This information will help constraining the model and testing its prediction on the importance of the strength of converging pathways. As simultaneous recordings of multiple cells (needed to test functional connectivity) remain challenging, this type of question is better addressed once a selective molecular marker has become available.

Most importantly, the knowledge of a molecular marker specific for anti-SWR cells would allow us to test whether their activity relates causally to the SWR incidence. The model predicts that SWR events can be triggered by inactivating a fraction of A cells (see Fig. 2.2D). Optogenetic inactivation could be a clean method to selectively target these cells in the slice and study the effects on the CA3 network.

In Section 4.3.1, I addressed the question of anti-SWR cells contribution to SWR generation by manipulating the activity of a single anti-SWR cell. Even though the results of the voltage clamp experiments are to date rather inconclusive (due to the low number of samples), they suggest that effects could possibly be observed in the slice recordings when cells are hyperpolarized. I predict that the same holds for experiments where anti-SWR cells are depolarized. The effect of single cell manipulation on SWR incidence might seem surprising, but earlier reports suggest that single neuron activity can be indeed detected at the network level [Brecht et al., 2004; Houweling and Brecht, 2008; Li et al., 2009; Kwan and Dan, 2012; Doron et al., 2014]. Specifically, Ellender et al. [2010] showed that the activation of a single PTI interneuron could increase SWR

incidence by up to 50%, and Bazelot et al. [2016] showed that driving a single pyramidal cell to spiking could trigger a SWR event (this effect, however, was not observed in Ellender et al. [2010]). The simulation results predict that an effect on SWR incidence could be observed if the number of anti-SWR cells is found to be small (around 50 CA3 cells in a slice, comparable to the model). If this is not the case, it would be important to refine the simulations by including the correct number of cells, or a non-random component in the connectivity of anti-SWR cells.

In the analysis presented in Sec. 4.3.1, I have focused on the SWR incidence as the preferred measure to detect changes in the network dynamics. However, it would be important to understand whether other properties (as SWR amplitude or duration, or ripple frequency) are also affected by changes in anti-SWR cells firing, both in experiments and in the model.

4.5 Methods

4.5.1 Experimental procedure

Animals. The housing and experimental use of animals was conducted in adherence to the guidelines of local authorities (Berlin state government, T0100/03) and to the German Animal Welfare Act and European Council Directive 2010/63/EU regarding the protection of animals used for experimental and other scientific purposes. Male mice between 4 and 12 weeks of age were used in experiments.

Slice preparation. Horizontal slices of ventral to mid-hippocampus were prepared as described before [Swaminathan et al., 2018]. Briefly, following isoflurane-anesthesia, brains were transferred to ice-cold sucrose-based aCSF containing (in mM): 87 NaCl, 2.5 KCl, 3 $MgCl_2$, 6 H_2O, 0.5 $CaCl_2$, 10 glucose, 50 sucrose, 1.25 NaH_2PO_4, and 26 $NaHCO_3$ (pH 7.4). Horizontal slices (400 μm) of ventral to mid-hippocampus were cut on a vibratome (VT1200S, Leica) and stored in an interface chamber perfused with aCSF containing (in mM): 119 NaCl, 2.5 KCl, 1.3 $MgCl_2$, 2.5 $CaCl_2$, 10 glucose, 1.25 NaH_2PO_4, and 26 $NaHCO_3$, at pH 7.4; osmolarity of 290 to 310 mosmol/l. The temperature was kept at $\sim$32° C, and slices were superfused at a rate of $\sim$1 ml/min. ACSF was equilibrated with carbogen (95% O_2, 5% CO_2). Slices recovered for $>$ 1.5 h after preparation.

Electrophysiological recordings. Combined LFP and single-cell ('loose-patch' or whole-cell) recordings were performed at 31° C-32° C in a submerged recording chamber perfused at 5-6 ml/min. For the recording of SWRs, aCSF-filled LFP pipettes (0.2-0.3 MΩ) were placed in the CA3 pyramidal cell layer. For 'loose-patch' recordings, an aCSF-filled pipette (2–5 MΩ) was used to establish a loose seal to the target cell for the extracellular detection of action potentials. This recording mode was first used to examine the temporal relation of the firing behavior of candidate cells with respect to LFP

SWRs. Only cells exhibiting the desired firing pattern, i.e., tonic firing outside SWR events, but decreased firing during SWRs, were pursued for further recordings. These putative anti-SWR cells were then recorded in the whole cell configuration with an intracellular solution containing (in mM): 120 K-gluconate, 10 HEPES, 10 KCl, 5 EGTA, 2 $MgSO_4$, 7 H_2O, 3 MgATP, 1 Na_2GTP, 5 phosphocreatine Na, and 5.4 biocytin (0.2%); pH adjusted to 7.4 with KOH. Current injection in the current clamp mode was used to further characterize the cells' firing- and intrinsic membrane properties. LFP signals were amplified 1,000-fold, and all data were filtered at 4 kHz (Bessel filter), and digitized at 20 kHz with 16-bit resolution using an A/D converter (BNC-2090 board, National Instruments, or Axon Digidata 1550A, Molecular Devices). Whole cell and extracellular recordings were performed using a Multiclamp 700A or 700B amplifier (Molecular Devices). Data were stored using Igor Pro (Wavemetrics) or pClamp (Molecular Devices).

Data analysis. SWR detection was performed in MATLAB (The MathWorks) as described before [Maier et al., 2009]. Time windows of 800 ms around the ripple peak of identified SWRs were cut out from the LFP and the corresponding 'loose-patch' or intracellular traces, and were baseline-corrected by subtracting the respective baseline means. Digital filtering was performed with a second-order Butterworth filter between 0.1 and 50 Hz. Spike times were detected using a threshold algorithm (4 standard deviations of the spike-free baseline). For each experiment, a histogram of the spike timings for all SWR events, aligned by ripple peak, were generated to examine the temporal relation of spiking and SWRs.

4.5.2 Simulations in the spiking network

To test the effect of modifying the firing of a single anti-SWR cell in the network at a time, I used the default spiking network displaying spontaneous SWRs introduced in Sec. 2.3. All network parameters are listed in Tables 2.1, 2.2, 2.3. First, I selected one anti-SWR cell and injected constant hyperpolarizing current ($I = -500$ pA), while monitoring the network activity for $T = 10$ minutes. I repeated the procedure once for each anti-SWR cell, for a total of 50 simulations. To test that each of the simulations was representative of the manipulation of a single cell, I randomly selected 3 anti-SWR cells and run 15 additional trials ($T = 10$ minutes each), during which the same cell was hyperpolarized. As only minor differences were observed in the SWR incidence across simulations in which the same cell was stimulated, I concluded that one simulation per cell is representative for the manipulation effect. To get a statistics of the control condition, I also run the network without any manipulation for 50 trials of $T = 10$ minutes. The procedure described here was replicated for the depolarized case, with injected current $I = 150$ pA. SWR incidence was calculated by low-pass filtering (Butterworth filter, second order) the instantaneous B population activity up to 5 Hz and detecting peaks larger than 20 Hz, at least 100 ms apart. This procedure is comparable to the more accurate method described in Sec. 2.8.4 and was here preferred because it reduces the memory load (otherwise needed to store the B input currents to the P cells).

4.5.3 Voltage clamp data analysis

Due to technical constraints, the analysis was performed in 6 of the 7 identified anti-SWR cells. In 4 recordings, the cells were recorded in the current clamp and voltage clamp configuration. In one of these recordings (id: 080219), the cell was first recorded in current clamp for 280 seconds and successively clamped at holding potentials $V = -60$ mV and $V = -70$ mV, each for a duration of 40 seconds. In another recording (id: 180319), the cell was in current clamp mode for 425 seconds before it was clamped at $V = -60$ mV and $V = -70$ mV for 125 seconds each. For the analysis, I merged the recordings were the cell was held at $V = -60$ mV and $V = -70$ mV. In this recording, current (100 pA) was injected to the cell in the current clamp mode to make it spike.

In three recordings (id: 230108, 210319 and 250319) only one voltage clamp value was tested. In the first set of recordings (id: 230108), the activity was recorded in current clamp for 220 seconds; additionally, the cell was clamped at $V = -60$ mV for 75 seconds. In the second set (id: 210319), the cell was recorded in current clamp for 275 seconds and then held at $V = -60$ for 150 seconds. Afterward, 150 seconds were recorded in the current clamp mode. For the analysis, the two CC periods were merged. In the recording with id: 250319, ten 20 seconds-long periods in current clamp were interleaved with periods (20 seconds, eleven periods in total) where hyperpolarizing current was injected in the cell so that its membrane potential was stable around $V = -65$ mV. For the analysis, this condition belongs to the voltage clamp mode, even though, technically, this is not a voltage clamp configuration. The interleaved approach has the advantage of distributing the possible time-dependent changes in slice health equally among the two conditions and should be considered the preferred paradigm in future experiments.

The results from one slice (id: 130219) were discarded because the SWR incidence was very low (< 0.15/s) and the clamping intervals were too short (60 seconds in total) to observe a sufficient number of spikes.

SWR events were detected by band-pass filtering the LFP traces in the range 2-20 Hz (Butterworth filter, second order) and detecting peaks above a given threshold (at least 100 ms apart). The threshold value in current and voltage clamp mode was adjusted to match the recording conditions and was confirmed by visual inspection.

5 | Outlook

In this last chapter, I summarize the contribution of this thesis to the broader field of memory research and discuss the main open questions that arose over the course of the project. Moreover, I emphasize alternative formulations of the model and the importance of additional work to test the model hypotheses.

5.1 Modeling choices and related work

In this work, I introduced a biologically realistic model to explain the generation of SWR events in the CA3 area of a rodent hippocampal slice. The model comprises three populations (pyramidal cells and two types of interneurons) and is based on a disinhibition mechanism that allows the pyramidal cells to spike during a SWR event.

Novel contribution. This model accounts for a number of experimental results that were not, to my knowledge, addressed in previous work. First, an important contribution of the model is that it explains the paradoxical effect of PV^+ cell stimulation [Schlingloff et al., 2014; Kohus et al., 2016] on the occurrence of SWRs. More in general, the recruitment of PV^+BCs for SWR generation in the model is in line with experiments showing an involvement of inhibitory neurons during the initial phase of a SWR [Ellender et al., 2010; Sasaki et al., 2014; Bazelot et al., 2016]. Furthermore, the dynamics of spontaneous SWRs reproduces the results by Kohus et al. [2016]; Chenkov [2017] on the existence of a strong correlation between SW amplitude and length of the previous (but not the next) IEI. Surprisingly, this feature is conserved in the spiking network across a number of conditions (additional depression and facilitation mechanisms, and gabazine modulation). I predict that this property mainly depends on the mechanism regulating the termination of SWRs. A valuable feature of the spiking network is that it does not rely on a strong recurrent connectivity among pyramidal cells, a feature that was widely used in previous SW models (see Sec. 1.2.4.1) but has recently been challenged by Guzman et al. [2016]. In the model presented in this thesis, I propose disinhibition as an alternative mechanism for SWR generation.

Disinhibition. The concept of disinhibition has attracted attention in the neuroscience community after several studies have shown that it is an ubiquitous feature of cortical circuits [Silberberg and Markram, 2007; Pfeffer et al., 2013; Letzkus et al., 2015; Karnani

et al., 2016; Pelkey et al., 2017]. From a mechanistic perspective, disinhibition has been used to successfully explain, among others, visual object-based attention, fear behavior in amygdala and prefrontal cortex, and sound tuning in auditory cortex [Van Der Velde and Kamps, 2001; Wolff et al., 2014; Takesian et al., 2018; Xu et al., 2019]. Hippocampal networks are also endowed with disinhibitory motifs, and it has been suggested that they could play an important role in establishing long-lasting memory traces in an object-recognition task [Donato et al., 2013]. However, the existence of the disinhibitory connection from PV^+BCs to anti-SWR cells has not been shown experimentally (the main obstacle being the yet unknown identity of anti-SWR cells), and its involvement in SWR generation remains an open question.

Termination of SWRs. In the context of my model, SWR events terminate because of a short-term synaptic depression mechanism that reduces the efficacy of the disinhibitory connection. In Section 2.5.1, I have explored the effect of alternative plastic changes in the network. However, other mechanisms could also support the termination of SWRs. For example, spike frequency adaptation is an additional mechanism that is commonly used in theoretical work as an alternative to synaptic depression (see e.g. [Shpiro et al., 2009; Levenstein et al., 2019]). The presence of spike frequency adaptation in the biological network remains to be explored. Alternatively, English et al. [2014] proposed that cellular hyperpolarization following a SWR event could induce a period where pyramidal cells are silent. The hyperpolarization could be the result of the activation of Ca^{2+}-dependent or ether-a-go-go (ERG) potassium currents [Zhang et al., 2006; Fano et al., 2012].

Ripple oscillations. In Section 2.3, I have shown that the SWR events generated by the spiking model are endowed with fast oscillations in the ripple range. Additional work is needed to study experimentally observed ripple features in the network (as intra-ripple-frequency accommodation and the phase of units firing during ripples [Ponomarenko et al., 2004; Nguyen, 2009; Csicsvari et al., 1999b; Stark et al., 2014]). In relation to the models of ripple generation discussed in Sec. 1.2.4.1, my approach suggests that the existence of recurrent inhibition between PV^+BCs and their connection to excitatory cells is sufficient to generate high-frequency oscillations.

Bistability. In the model, the main mechanism accounting for the generation of SWRs is bistability, i.e., the coexistence of SWR and non-SWR states. Practically, the interneurons compete for regulating the firing activity of pyramidal cells. In a perfect bistable configuration, SWRs can only be induced by current injection, but do not arise spontaneously. By destabilizing the non-SWR state, small fluctuations in the network activity can suffice to start an event. This type of inhibitory networks, where noisy behavior and adaptive mechanisms coexist, has been extensively studied [Moreno-Bote et al., 2007; Shpiro et al., 2007; Curtu et al., 2008; Shpiro et al., 2009; Jercog et al., 2017; Levenstein et al., 2019], but mostly in the scenario with only one or two populations. These models

inspired the definition of SWRs as events arising from a bistable network that is perturbed by noise (in my case, the noise is intrinsic in the spiking network). Nevertheless, it remains to be explored if bistability is present in the biological network. The model predicts that, once a SWR event has been initiated, reducing the synaptic depression at the disinhibitory connection would prolong the duration of the SWR. However, I expect that, in the biological system, additional mechanisms are in place to prevent a overlong SWR-like activity, which could damage the network. For example, alternative plasticity mechanisms, as the pyramidal-to-anti-SWR cells facilitation proposed in Sec. 2.5.1, could contribute to restore the default non-SWR state.

Relation to *in vivo* SWRs. To tune the model to biologically realistic results, I largely relied on *in vitro* data (Sec. 1.2.3). Nevertheless, I expect that the proposed model also captures the most relevant features underlying SWR occurrence *in vivo* (see Sec. 1.2.1 for a comparison between *in vivo* and *in vitro* SWRs). This hypothesis could be tested by scaling up the number of cells modeled in the network and verifying that the main results presented in Chapter 2 still hold true. A crucial point to be addressed is how to scale the connectivity among populations to account for the increased network size while keeping the network noise large enough to trigger spontaneous SWRs. Additional features characteristic of an *in vivo* setup (as e.g. cortical inputs to select which cells initiate an event) could further expand the core model presented in this thesis (also see Sec. 5.2).

Effect of PV$^+$ cell activation. The finding on PV$^+$ cell activation triggering SWR events [Schlingloff et al., 2014; Kohus et al., 2016] was the main motivation to include three populations in the network. In reality, I restricted myself to the modeling of PV$^+$ basket cells, a subset of PV$^+$ cells whose involvement in SWRs has been intensively studied [Klausberger and Somogyi, 2008; Hájos et al., 2013; Kohus et al., 2016]. However, I cannot exclude the hypothesis that other types of PV$^+$ cells (e.g., bistratified or axo-axonic cells) also play an important role in SWR generation. This means, on the one hand, that it remains to be tested whether the results of Schlingloff et al. [2014] hold true when PV$^+$ basket cells are selectively activated. On the other hand, if other PV$^+$ populations, which behave differently from PV$^+$BCs, are found to play an important role in SWR generation, the model needs to be extended to include these classes of cells.

Alternative model architectures. Another key point that needs to be addressed more in detail is whether alternatives to the three-population model exist, which can reproduce the results of Schlingloff et al. [2014]; Kohus et al. [2016]. At this point, I cannot exclude that alternative network architectures that are not based on the disinhibition mechanism could also generate a SWR event upon PV$^+$ cells activation. For example, Cobb et al. [1995]; Stark et al. [2013] have shown that inhibitory inputs to depolarized

excitatory cells can lead to a rebound (brief depolarization) which follows the initial hyperpolarization. If this behavior happened at multiple neurons simultaneously, a population burst (SW) could emerge. However, this hypothesis is not supported by experiments showing that SWRs can be initiated with short latencies (< 10 ms in [Schlingloff et al., 2014; Stark et al., 2014; Bazelot et al., 2016], and Sec. 1.2.3). More in general, the results on the short latency between stimulus and SWR onset hint at a mechanism involving only fast activating components, such as mono- or disynaptic connections between PV^+ and pyramidal cells.

It is however important to mention the results of Ellender et al. [2010], who showed *in vitro* that the activation of perisomatic-targeting interneurons (PTI, a group that includes basket cells) elicits SWRs with latencies up to 1.5 seconds. In these experiments, the SWR incidence decreased during the (500 ms-long) stimulation time, an effect that cannot be reproduced in my model. Due to this effect, and the long latency preceding SWR onset, the generating mechanism might in this case be different from the one hypothesized by my model. Potentially linked to these results, Hulse et al. [2016] showed that a 1 second-long depolarization preceding a SWR event can be observed in a subset of CA1 pyramidal cells. Although differences might exist between the *in vitro* and *in vivo* setup, both findings hints at the existence of a slow mechanism linked to the initiation of a SWR. It remains to be explored, both in experiments and in the model, how the slow build-up of activity relates to the correlation structure reported in Kohus et al. [2016]; Jiang et al. [2018], and, more in general, to the mechanisms regulating the length of SWR inter-event-intervals.

Model generality. Finally, the analysis of the mean-field network demonstrates that the results presented in this thesis are largely independent of the exact model formulation (population size, intrinsic neuronal properties, connection strengths, population firing rates, etc.). Thus, the model is able to describe general principles of a wide range of bistable disinhibitory networks, and I predict that the spiking network can be adapted as more experimental results on the identity of anti-SWR cells become available.

5.2 Future directions and the link to memory consolidation

Despite the fact that the proposed model can explain a number of relevant experimental features, there are broader questions that the model inevitably fails to address. In this last section, I briefly review some directions in which this work could be extended.

Pyramidal cells do not form a homogeneous class. First, a recent paper by Hunt et al. [2018] convincingly shows that the population of CA3 pyramidal cells is not homogeneous, but can be divided into two groups of preferentially regular spiking and bursting neurons. These firing patterns seem to correlate with distinct electrophysiological properties, cell morphologies (thorny and athorny, respectively), incoming connectivity from the dentate gyrus (DG), cholinergic modulation, and cell-type specific activity patterns

during SWRs. Specifically, athorny cells contribute with burst firing to the building up of activity characteristic of SWRs, whereas regular spiking cells fire (mostly single spikes) around the peak of the SW. Thus, the authors suggest that athorny cells could play an important role in SWR generation. Even though a causal link is still missing, it would be interesting to extend the model to include a population of bursting pyramidal cells, once more is known about their embedding in the local circuit.

Linking the SWR model to other cortico-hippocampal rhythms. In Section 2.7, I proposed to incorporate inhomogeneous populations in the model. The aim of this extension is twofold. On the one hand, it could be sufficient to explain the effect of single cell stimulation reported in Bazelot et al. [2016]. On the other hand, it could provide a link with the existence of sequence reactivation during SWRs [Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996]. In the model, the pyramidal cell population could be organized in cell assemblies (with strong internal connectivity), and the connectivity between assemblies could be tuned to yield sequential activation in correspondence of SWRs. This approach has been investigated by Chenkov et al. [2017], but it is unclear whether the results can be replicated in the three-population scenario. If successful, the approach would yield a unified model of hippocampal activity during SWRs (similarly to the attempts in Jahnke et al. [2015]; Malerba and Bazhenov [2019]). This approach is in line with the notion of SWRs being windows of opportunity for the occurrence of memory-related sequence replay [Lee and Wilson, 2002; O'Neill et al., 2008].

What could be alternative roles for SWRs? One prominent hypothesis is that SWRs are means of cortico-hippocampal communications during offline periods [Isomura et al., 2006; Ji and Wilson, 2007]. In this scenario, hippocampal SWR activity is relevant in combination with cortical slow oscillations, to transfer the encoded memories from the hippocampus to neocortical areas. At the cellular level, slow oscillations are related to the alternation of UP / DOWN periods (see Sec. 1.1.4.2). A possible extension of this work could include the modeling of cortico-hippocampal communication, where SWRs could be induced by cortical activity, and vice versa. Intriguingly, a recent model hypothesized that the generating mechanisms of SWRs and UP / DOWN states are identical, except for the relative balance of active and inactive states [Levenstein et al., 2019]. According to this hypothesis, SWRs can be interpreted as brief interruptions of the more stable non-SWR state, whereas during slow oscillations, stable UP periods are interrupted by transient DOWN states. The conceptual formulation of the model by Levenstein et al. [2019] hinders a direct comparison to known biological results, but the underlying idea of regarding a SWR as a brief disruption of a resting stable state is in line with my approach. Once more is known about the microcircuit responsible for UP / DOWN states generation, it could be possible to assess if my model can be extended to explain the generation of UP / DOWN transitions. In particular, the involvement of PV^+ interneurons in the generation of UP / DOWN states needs to be clarified [Kuki et al., 2015; Neske et al., 2015; Funk et al., 2017; Zucca et al., 2017; Niethard et al., 2018],

and cells that preferentially fire during the DOWN states (which should correspond to the anti-SWR cells in my model) have not been identified so far [Neske et al., 2015].

A provocative hypothesis suggests that LFP signals are an epiphenomenon of cells' synaptic interaction and spiking activity (see Herreras [2016]; Luo and Guan [2018] for recent reviews on this matter). Even though I cannot exclude that this is indeed the case, it seems improbable that the most synchronous event in the mammalian brain [Buzsáki, 2015] is just a side effect of neuronal activity. An open possibility is, however, that SWRs (and LFP signals more in general) influence spiking activity by means of the so called ephaptic coupling. Experimental and theoretical work [Chan and Nicholson, 1986; Ozen et al., 2010; Anastassiou et al., 2010, 2011] has indeed shown that synchronous events like SWRs generate an electric field that can be large enough to influence the spike threshold and timing of spikes, suggesting that SWRs could play a role in synchronizing neuronal activity. These hypotheses on the putative role of SWRs are by no means mutually exclusive, and are likely to synergistically contribute to different aspects of offline hippocampal function.

As I reviewed in Section 1.1.3.1, the same hippocampal regions involved in SWR generation during offline periods of slow-wave sleep and awake rest are characterized by theta and gamma oscillations during active exploration and REM sleep. It is unclear how these regions can generate both types of rhythms, and what regulates the switch between the two modes. The evidence on the importance of neuromodulation (Sec. 1.2.2) and the fact that SWRs, but not theta oscillations, happen spontaneously in the slice (Sec. 1.2.1) suggests that these questions should be tackled with an *in vivo* approach. I believe, however, that modeling work could contribute to answering conceptual questions about the mechanisms allowing the switch between SWRs and theta-gamma oscillations. In particular, future work could address the possibility of integrating the SWRs model with existing models of gamma oscillations (see Bartos et al. [2007] for a review) and extending this framework to incorporate neuromodulatory effects [Hasselmo, 1999].

The role of anti-SWR cells. The main prediction of the model is the existence of a group of cells that are anti-modulated with respect to SWRs. Despite the encouraging results presented in Chapter 4, the identity of the anti-SWR cells remains to be clarified. Specifically, it would be important to classify these cells in terms of morphological and electrophysiological properties, and to study their immunoreactivity to different molecular markers as well as their embedding in the hippocampal network. The prospect of identifying a molecular marker selective for this cell type would open the exciting opportunity of testing the model prediction on the inactivation of anti-SWR cells triggering a SWR events. If the hypothesis holds true, the modulation of the activity of anti-SWR cells could provide an effective method to regulate the occurrence of SWRs, and, in the long run, to interfere with the process of memory consolidation.

Bibliography

Abbott, L. F. (1994). Decoding neuronal firing and modeling neural networks. *Quarterly reviews of biophysics*, 27(3):291–331.

Abraham, W. C., Gustafsson, B., and Wigström, H. (1986). Single high strength afferent volleys can produce long-term potentiation in the hippocampus in vitro. *Neuroscience Letters*, 70(2):217–222.

Abraham, W. C. and Robins, A. (2005). Memory retention - The synaptic stability versus plasticity dilemma. *Trends in Neurosciences*, 28(2):73–78.

Abraham, W. C. and Williams, J. M. (2003). Properties and mechanisms of LTP maintenance. *The Neuroscientist*, 9(6):463–474.

Achermann, P. and Borbély, A. (1997). Low frequency (< 1Hz) oscillations in the human sleep electroencephalogram. *Neuroscience*, 81(1):213–222.

Adrian, E. D. (1942). Olfactory reactions in the brain of the hedgehog. *Journal of Physiology*, 100(4):459–473.

Ali, A. B. and Thomson, A. M. (1998). Facilitating pyramid to horizontal oriens-alveus interneurone inputs: Dual intracellular recordings in slices of rat hippocampus. *Journal of Physiology*, 507(1):185–199.

Alonso, A. and Klink, R. (1993). Differential electroresponsiveness of stellate and pyramidal-like cells of medial entorhinal cortex layer II. *Journal of Neurophysiology*, 70(1):128–143.

Amaral, D. G. and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience*, 31(3):571–591.

Amzica, F. and Steriade, M. (1997). The K-complex: its slow (<1 Hz) rhythmicity and relation to delta waves. *Neurology*, 49(4):952–959.

Anastassiou, C. A., Montgomery, S. M., Barahona, M., Buzsaki, G., and Koch, C. (2010). The effect of spatially inhomogeneous extracellular electric fields on neurons. *Journal of Neuroscience*, 30(5):1925–1936.

Anastassiou, C. A., Perin, R., Markram, H., and Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nature Neuroscience*, 14(2):217–23.

Arnolds, D., Lopes Da Silva, F., Aitink, J., Kamp, A., and Boeijinga, P. (1980). The spectral properties of hippocampal EEG related to behaviour in man. *Electroencephalography and Clinical Neurophysiology*, 50(3-4):324–328.

Augustin, M., Ladenbauer, J., Baumann, F., and Obermayer, K. (2017). Low-dimensional spike rate models derived from networks of adaptive integrate-and-fire neurons: Comparison and implementation. *PLoS Computational Biology*, 13(6):1–46.

Axmacher, N., Elger, C. E., and Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain*, 131(7):1806–1817.

Azizi, A. H., Wiskott, L., and Cheng, S. (2013). A computational model for preplay in the hippocampus. *Frontiers in Computational Neuroscience*, 7(161):1–15.

Bähner, F., Weiss, E. K., Birke, G., Maier, N., Schmitz, D., Rudolph, U., Frotscher, M., Traub, R. D., Both, M., and Draguhn, A. (2011). Cellular correlate of assembly formation in oscillating hippocampal networks in vitro. *Proceedings of the National Academy of Sciences*, 108(35):E607–E616.

Bartos, M., Vida, I., Frotscher, M., Meyer, A., Monyer, H., Geiger, R. P., and Jonas, P. (2002). Fast synaptic inhibition promotes synchronized gamma oscillations in hippocampal interneuron networks. *Proceedings of the National Academy of Sciences*, 99(20):13222–13227.

Bartos, M., Vida, I., and Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nature Reviews Neuroscience*, 8(1):45–56.

Battaglia, F. P., Sutherland, G. R., and McNaughton, B. L. (2004). Hippocampal sharp wave bursts coincide with neocortical "up-state" transitions. *Learning and Memory*, 11(6):697–704.

Bazelot, M., Teleńczuk, M. T., and Miles, R. (2016). Single CA3 pyramidal cells trigger sharp waves in vitro by exciting interneurones. *Journal of Physiology*, 594(10):2565–2577.

Behrens, C. J., van den Boom, L. P., de Hoz, L., Friedman, A., and Heinemann, U. (2005). Induction of sharp wave-ripple complexes in vitro and reorganization of hippocampal networks. *Nature Neuroscience*, 8(11):1560–1567.

Behrens, C. J., Van Den Boom, L. P., and Heinemann, U. (2007). Effects of the GABAA receptor antagonists bicuculline and gabazine on stimulus-induced sharp wave-ripple complexes in adult rat hippocampus in vitro. *European Journal of Neuroscience*, 25(7):2170–2181.

Beyeler, A., Retailleau, A., Molter, C., Mehidi, A., Szabadics, J., and Leinekugel, X. (2013). Recruitment of perisomatic inhibition during spontaneous hippocampal activity in vitro. *PLoS ONE*, 8(6):e66509.

Bezaire, M. J. and Soltesz, I. (2013). Quantitative assessment of CA1 local circuits: Knowledge base for interneuron-pyramidal cell connectivity. *Hippocampus*, 23(9):751–785.

Bhatia, A., Moza, S., and Bhalla, U. S. (2019). Precise excitation-inhibition balance controls gain and timing in the hippocampus. *eLife*, 8:e43415.

Bi, G.-q. and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472.

Binder, S., Baier, P. C., Mölle, M., Inostroza, M., Born, J., and Marshall, L. (2012). Sleep enhances memory consolidation in the hippocampus-dependent object-place recognition task in rats. *Neurobiology of Learning and Memory*, 97(2):213–219.

Binder, S., Rawohl, J., Born, J., and Marshall, L. (2014). Transcranial slow oscillation stimulation during NREM sleep enhances acquisition of the radial maze task and modulates cortical network activity in rats. *Frontiers in Behavioral Neuroscience*, 7(220):1–10.

Bland, B. H. (1986). The physiology and pharmacology of hippocampal formation theta rhythms. *Progress in Neurobiology*, 26(1):1–54.

Bliss, T. V. P. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *Journal*

of Physiology, 232(2):357–374.

Böhm, C., Pangalos, M., Schmitz, D., and Winterer, J. (2015). Serotonin attenuates feedback excitation onto O-LM interneurons. *Cerebral Cortex*, 25(11):4572–4583.

Booker, S. A. and Vida, I. (2018). Morphological diversity and connectivity of hippocampal interneurons. *Cell and Tissue Research*, 373(3):619–641.

Borg-Graham, L. J., Monier, C., and Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393(6683):369–373.

Both, M., Bähner, F., Von Bohlen und Halbach, O., and Draguhn, A. (2008). Propagation of specific network patterns through the mouse hippocampus. *Hippocampus*, 18(9):899–908.

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., and Buzsáki, G. (1995a). Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *Journal of Neuroscience*, 15(1 Pt 1):47–60.

Bragin, A., Jando, G., Nadasdy, Z., van Landeghem, M., and Buzsáki, G. (1995b). Dentate EEG spikes and associated interneuronal population bursts in the hippocampal hilar region of the rat. *Journal of Neurophysiology*, 73(4):1691–1705.

Branchereau, P., Van Bockstaele, E. J., Chan, J., and Pickel, V. M. (1996). Pyramidal neurons in rat prefrontal cortex show a complex synaptic response to single electrical stimulation of the locus coeruleus region: Evidence for antidromic activation and GABAergic inhibition using in vivo intracellular recording and electron micr. *Synapse*, 22(4):313–331.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3):340–352.

Brecht, M., Schneider, M., Sakmann, B., and Margie, T. W. (2004). Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, 427(6976):704–710.

Bressler, S. L. and Freeman, W. J. (2003). Frequency analysis of olfactory system EEG in cat, rabbit, and rat. *Electroencephalography and Clinical Neurophysiology*, 50(1-2):19–24.

Brunel, N. (2000a). Dynamics of sparsely connected networls of excitatory and inhibitory neurons. *Journal of Computational Neuroscience*, 8(3):183–208.

Brunel, N. (2000b). Persistent activity and the single-cell frequency-current curve in a cortical network model. *Network: Computation in Neural Systems*, 11(4):261–280.

Brunel, N. and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*, 90(1):415–30.

Buhl, D. and Buzsáki, G. (2005). Developmental emergence of hippocampal fast-field "ripple" oscillations in the behvaing rat pups. *Neuroscience*, 134(4):1423–1430.

Burwell, R. D. and Witter, M. P. (2002). Basic anatomy of the parahippocampal region in rats and monkeys. In Witter, M. P. and Wouterlood, F. G., editors, *The Parahippocampal Region, Organization and Role in Cognitive Functions*, pages 35– 60. Oxford University Press, Oxford, UK.

Buzsáki, G. (1986). Hippocampal sharp waves: their origin and significance. *Brain research*, 398:242–252.

Buzsáki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience*, 31(3):551–70.

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33(3):325–40.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, 25(10):1073–1188.

Buzsáki, G., Haas, H. L., and Anderson, E. G. (1987). Long-term potentiation induced by physiologically relevant stimulus patterns. *Brain Research*, 435(1-2):331–333.

Buzsáki, G., Horváth, Z., Urioste, R., Hetke, J., and Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025–1027.

Buzsáki, G., Leung, L. W. S., and Vanderwolf, C. H. (1983). Cellular bases of hippocampal EEG in the behaving rat. *Brain Research Reviews*, 6(2):139–171.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628.

Canto, C. B., Wouterlood, F. G., and Witter, M. P. (2008). What does the anatomical organization of the entorhinal cortex tell us? *Neural Plasticity*, 2008:1–18.

Castro-Alamancos, M. A. and Connors, B. W. (2002). Distinct forms of short-term plasticity at excitatory synapses of hippocampus and neocortex. *Proceedings of the National Academy of Sciences*, 94(8):4161–4166.

Chan, C. Y. and Nicholson, C. (1986). Modulation by applied electric fields of Purkinje and stellate cell activity in the isolated turtle cerebellum. *Journal of Physiology*, 371(1):89–114.

Cheng, S. and Frank, L. M. (2008). New experiences enhance coordinated neural activity in the hippocampus. *Neuron*, 57(2):303–313.

Chenkov, N. (2017). *Network mechanisms underlying sharp-wave ripples and memory replay*. PhD thesis, Humboldt-Universität zu Berlin, Lebenswissenschaftliche Fakultät.

Chenkov, N., Sprekeler, H., and Kempter, R. (2017). Memory replay in balanced recurrent networks. *PLoS Computational Biology*, 13(1):e1005359.

Chizhov, A. V., Sanchez-Aguilera, A., Rodrigues, S., and De La Prida, L. M. (2015). Simplest relationship between local field potential and intracellular signals in layered neural tissue. *Physical Review E*, 92(6):1–9.

Chrobak, J. J. and Buzsáki, G. (1994). Selective activation of deep layer (V-VI) retrohippocampal cortical neurons during hippocampal sharp waves in the behaving rat. *Journal of Neuroscience*, 14(10):6160–6170.

Clark, R. E. and Squire, L. R. (2013). Similarity in form and function of the hippocampus in rodents, monkeys, and humans. *Proceedings of the National Academy of Sciences*, 110(suppl. 2):10365–10370.

Clemens, Z., Mölle, M., Eross, L., Barsi, P., Halász, P., and Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain*, 130(11):2868–2878.

Clemens, Z., Mölle, M., Eross, L., Jakus, R., Rásonyi, G., Halász, P., and Born, J. (2011). Fine-tuned coupling between human parahippocampal ripples and sleep spindles. *European Journal of Neuroscience*, 33(3):511–520.

Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O., and Somogyi, P. (1995). Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature*,

378(6552):75–78.

Cobb, S. R., Halasy, K., Vida, I., Nyiri, G., Tamás, G., Buhl, E. H., and Somogyi, P. (1997). Synaptic effects of identified interneurons innervating both interneurons and pyramidal cells in the rat hippocampus. *Neuroscience*, 79(3):629–648.

Colgin, L. L. (2015). Theta-gamma coupling in the entorhinal-hippocampal system. *Current Opinion in Neurobiology*, 31:45–50.

Colgin, L. L., Kubota, D., Jia, Y., Rex, C. S., and Lynch, G. (2004). Long-term potentiation is impaired in rat hippocampal slices that produce spontaneous sharp waves. *Journal of Physiology*, 558(3):953–961.

Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X.-J. (2003). Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. *Journal of Neurophysiology*, 89(5):2707–2725.

Csicsvari, J., Hirase, H., Czurkó, A., Mamiya, A., and Buzsáki, G. (1999a). Fast network oscillations in the hippocampal CA1 region of the behaving rat. *Journal of Neuroscience*, 19(RC20):1–4.

Csicsvari, J., Hirase, H., Czurkó, A., Mamiya, A., and Buzsáki, G. (1999b). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *Journal of Neuroscience*, 19(1):274–287.

Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble patterns of hippocampal CA3-CA1 neurons during sharp wave-associated population events. *Neuron*, 28(2):585–594.

Csicsvari, J., O'Neill, J., Allen, K., and Senior, T. (2007). Place-selective firing contributes to the reverse-order reactivation of CA1 pyramidal cells during sharp waves in open-field exploration. *European Journal of Neuroscience*, 26(3):704–716.

Curtu, R., Shpiro, A., Rubin, N., and Rinzel, J. (2008). Mechanisms for frequency control in neuronal competition models. *SIAM journal on applied dynamical systems*, 7(2):609–649.

Cutsuridis, V., Cobb, S., and Graham, B. P. (2010). Encoding and retrieval in a model of the hippocampal CA1 microcircuit. *Hippocampus*, 20(3):423–446.

Dave, A. S. and Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*, 290(5492):812–816.

Davidson, T. J., Kloosterman, F., and Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, 63(4):497–507.

Davoudi, H. and Foster, D. J. (2019). Acute silencing of hippocampal CA3 reveals a necessary role in place field responses. *Nature Neuroscience*, 22(3):337–342.

Dayan, P. and Abbott, L. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press.

De Gennaro, L. and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep medicine reviews*, 7(5):423–40.

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neuroscience*, 4(9):739–751.

Di Volo, M., Romagnoni, A., Capone, C., and Destexhe, A. (2019). Biologically Realistic Mean-Field Models of Conductance-Based Networks of Spiking Neurons with Adaptation. *Neural Computation*, 31(4):653–680.

Diba, K. and Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10):1241–1242.

Diekelmann, S. and Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2):114–126.

Domnisoru, C., Kinkhabwala, A. A., and Tank, D. W. (2013). Membrane potential dynamics of grid cells. *Nature*, 495(7440):199–204.

Donato, F., Rompani, S. B., and Caroni, P. (2013). Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature*, 504(7479):272–276.

Donoso, J. R. (2016). *The role of interneuronal networks in hippocampal ripple oscillations*. PhD thesis, Humboldt-Universität zu Berlin, Lebenswissenschaftliche Fakultät.

Donoso, J. R., Schmitz, D., Maier, N., and Kempter, R. (2018). Hippocampal ripple oscillations and inhibition-first network models: Frequency dynamics and response to GABA modulators. *Journal of Neuroscience*, 38(12):0188–17.

Doron, G., von Heimendahl, M., Schlattmann, P., Houweling, A. R., and Brecht, M. (2014). Spiking irregularity and frequency modulate the behavioral report of single-neuron stimulation. *Neuron*, 81(3):653–663.

Dougherty, K. A., Islam, T., and Johnston, D. (2012). Intrinsic excitability of CA1 pyramidal neurones from the rat dorsal and ventral hippocampus. *Journal of Physiology*, 590(22):5707–5722.

Dragoi, G. and Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401.

Draguhn, A., Traub, R. D., Schmitz, D., and Jefferys, J. G. R. (1998). Electrical coupling underlies high-frequency oscillations in the hippocampus in vitro. *Nature*, 394(6689):189–192.

Duarte, M. (2015). Notes on Scientific Computing for Biomechanics and Motor Control. https://github.com/demotu/BMC.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, 55(1):51–86.

Dugas, C., Bengio, Y., and Belisle, F. (2001). Incorporating second-order functional knowledge for better option pricing. In *NIPS 13*, pages 472–478.

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., and Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature Neuroscience*, 13(8):995–1002.

Dur-e Ahmad, M., Nicola, W., Campbell, S. A., and Skinner, F. K. (2012). Network bursting using experimentally constrained single compartment CA3 hippocampal neuron models with adaptation. *Journal of Computational Neuroscience*, 33(1):21–40.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., and Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, 94(6):4269–4280.

Ego-Stengel, V. and Wilson, M. A. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, 20:1–10.

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1(1):41–50.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785.

Ellender, T. J., Nissen, W., Colgin, L. L., Mann, E. O., and Paulsen, O. (2010). Priming of hippocampal population bursts by individual perisomatic-targeting interneurons. *Journal of Neuroscience*, 30(17):5979–91.

English, D. F., McKenzie, S., Evans, T., Kim, K., Yoon, E., and Buzsáki, G. (2017). Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks. *Neuron*, 96(2):505–520.

English, D. F., Peyrache, A., Stark, E., Roux, L., Vallentin, D., Long, M. A., and Buzsáki, G. (2014). Excitation and inhibition compete to control spiking during hippocampal ripples: Intracellular study in behaving mice. *Journal of Neuroscience*, 34(49):16509–16517.

Ermentrout, B. (1994). Reduction of conductance-based models with slow synapses to neural nets. *Neural computation*, 695(6):679–695.

Eschenko, O., Ramadan, W., Mölle, M., Born, J., and Sara, S. J. (2008). Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning & memory*, 15(4):222–228.

Euston, D. R., Tatsuno, M., and McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, 318(5853):1147–1150.

Fano, S., Çalişkan, G., and Heinemann, U. (2012). Differential effects of blockade of ERG channels on gamma oscillations and excitability in rat hippocampal slices. *European Journal of Neuroscience*, 36(12):3628–3635.

Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex. *Annual Review of Neuroscience*, 32(1):33–55.

Fellous, J. M., Rudolph, M., Destexhe, A., and Sejnowski, T. J. (2003). Synaptic background noise controls the input/output characteristics of single cells in an in vitro model of in vivo activity. *Neuroscience*, 122(3):811–829.

Foster, D. J. and Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683.

Frankland, P. W. and Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2):119–130.

Freund, T. F. and Buzsáki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, 6(4):347–470.

Fuentealba, P., Tomioka, R., Dalezios, Y., Márton, L. F., Studer, M., Rockland, K., Klausberger, T., and Somogyi, P. (2008). Rhythmically active enkephalin-expressing GABAergic cells in the CA1 area of the hippocampus project to the subiculum and preferentially innervate interneurons. *Journal of Neuroscience*, 28(40):10017–10022.

Funk, C. M., Peelman, K., Bellesi, M., Marshall, W., Cirelli, C., and Tononi, G. (2017). Role of somatostatin-positive cortical interneurons in the generation of sleep slow waves. *Journal of Neuroscience*, 37(38):9132–9148.

Fusi, S. and Mattia, M. (1999). Collective behavior of networks with linear (VLSI) integrate-and-fire neurons. *Neural Computation*, 11(3):633–652.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, 49(1):87–115.

Galarreta, M. and Hestrin, S. (1998). Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neuroscience*, 1(7):587–594.

Geiger, J. R., Melcher, T., Koh, D. S., Sakmann, B., Seeburg, P. H., Jonas, P., and Monyer, H. (1995). Relative abundance of subunit mRNAs determines gating and Ca 2+ permeability of AMPA receptors in principal neurons and interneurons in rat CNS. *Neuron*, 15(1):193–204.

Geisler, C., Brunel, N., and Wang, X.-J. (2005). Contributions of intrinsic membrane dynamics to fast network oscillations with irregular neuronal discharges. *Journal of Neurophysiology*, 94(6):4344–4361.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., and Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, 12(10):1222–1223.

Girardeau, G. and Cei, A. (2014). Learning-induced plasticity regulates hippocampal sharp wave-ripple drive. *Journal of Neuroscience*, 34(15):5176–5183.

Girardeau, G. and Zugaro, M. (2011). Hippocampal ripples and memory consolidation. *Current Opinion in Neurobiology*, 21(3):452–459.

Giri, B., Miyawaki, H., Mizuseki, K., Cheng, S., and Diba, K. (2018). Hippocampal reactivation extends for several hours following novel experience. *Journal of Neuroscience*, 39(5):866–875.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, pages 315–323.

Goodman, D. F. and Brette, R. (2009). The brian simulator. *Frontiers in Neuroscience*, 3(2):192–197.

Govindarajan, A., Kelleher, R. J., and Tonegawa, S. (2006). A clustered plasticity model of long-term memory engrams. *Nature Reviews Neuroscience*, 7(7):575–583.

Gray, C. M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86(5):1698–1702.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., and Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, 65(5):695–705.

Guzman, S. J., Schlogl, A., Frotscher, M., and Jonas, P. (2016). Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science*, 353(6304):1117–1123.

Ha, G. E. and Cheong, E. (2017). Spike frequency adaptation in neurons of the central nervous system. *Experimental Neurobiology*, 26(4):179.

Hahnloser, R. H., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902):65–70.

Hájos, N., Ellender, T. J., Zemankovics, R., Mann, E. O., Exley, R., Cragg, S. J., Freund, T. F., and Paulsen, O. (2009). Maintaining network activity in submerged hippocampal slices: Importance of oxygen supply. *European Journal of Neuroscience*, 29(1986):319–327.

Hájos, N., Karlócai, M. R., Németh, B., Ulbert, I., Monyer, H., Szabó, G., Erdélyi, F., Freund, T. F., and Gulyás, A. I. (2013). Input-output features of anatomically identified CA3 neurons during hippocampal sharp wave/ripple oscillation in vitro. *Journal of Neuroscience*, 33(28):11677–91.

Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, 12(9):509–523.

Hartse, K. M., Eisenhart, S. F., Bergmann, B. M., and Rechtschaffen, A. (1979). Ventral hippocampus spikes during sleep, wakefulness, and arousal in the cat. *Sleep*, 1(3):231–246.

Hasselmo, M. E. (1995). Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behavioural Brain Research*, 67(1):1–27.

Hasselmo, M. E. (1999). Neuromodulation: Acetylcholine and memory consolidation. *Trends in Cognitive Sciences*, 3(9):351–359.

Hasselmo, M. E. and Schnell, E. (1994). Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. *Journal of Neuroscience*, 14(6):3898–914.

Hasselmo, M. E., Schnell, E., and Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience*, 15(7):5249–62.

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. John Wiley and Sons, New York.

Herreras, O. (2016). Local field potentials: myths and misunderstandings. *Frontiers in Neural Circuits*, 10:101.

Hill, A. V. (1936). Excitation and accommodation in nerve. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, 119(814):305–355.

Hirase, H., Leinekugel, X., Czurko, A., Csicsvari, J., and Buzsáki, G. (2001). Firing rates of hippocampal neurons are preserved during subsequent sleep episodes and modified by novel awake experience. *Proceedings of the National Academy of Sciences*, 98(16):9386–9390.

Hirsch, J. A., Alonso, J.-M., Reid, R. C., and Martinez, L. M. (2018). Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18(22):9517–9528.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117(4):500–544.

Hofer, K. T., Kandrács, Á., Ulbert, I., and Pál, I. (2015). The hippocampal CA3 region can generate two distinct types of sharp wave-ripple complexes, in vitro. *Hippocampus*, 25(2):169–186.

Hoffman, K. L. and McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, 297(5589):2070–2073.

Holcman, D. and Tsodyks, M. (2006). The emergence of up and down states in cortical networks. *PLoS Computational Biology*, 2(3):174–181.

Hollnagel, J. O., Maslarova, A., ul Haq, R., and Heinemann, U. (2014). GABAB receptor dependent modulation of sharp wave-ripple complexes in the rat hippocampus in vitro. *Neuroscience Letters*, 574:15–20.

Holzbecher, A. and Kempter, R. (2018). Interneuronal gap junctions increase synchrony and

robustness of hippocampal ripple oscillations. *European Journal of Neuroscience*, 48(12):3446–3465.

Houweling, A. R. and Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451(7174):65–68.

Huang, E. P. (1998). Synaptic plasticity: Going through phases with LTP. *Current Biology*, 8(10):R350–R352.

Hulse, B. K., Moreaux, L. C., Lubenov, E. V., and Siapas, A. G. (2016). Membrane potential dynamics of CA1 pyramidal neurons during hippocampal ripples in awake mice. *Neuron*, 89(4):800–813.

Hunt, D. L., Linaro, D., Si, B., Romani, S., and Spruston, N. (2018). A novel pyramidal cell type promotes sharp-wave synchronization in the hippocampus. *Nature Neuroscience*, 21:985–995.

Hwaun, E. and Colgin, L. L. (2019). CA3 place cells that represent a novel waking experience are preferentially reactivated during sharp wave-ripples in subsequent sleep. *Hippocampus*, pages 1–18.

Ishizuka, N., Weber, J., and Amaral, D. G. (1990). Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *Journal of Comparative Neurology*, 295(4):580–623.

Isomura, Y., Sirota, A., Özen, S., Montgomery, S., Mizuseki, K., Henze, D. A., and Buzsáki, G. (2006). Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. *Neuron*, 52(5):871–882.

Jackson, J. C., Johnson, A., and Redish, A. D. (2006). Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience. *Journal of Neuroscience*, 26(48):12415–12426.

Jadhav, S. P., Kemere, C., German, P. W., and Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336(6087):1454–1458.

Jahnke, S., Timme, M., and Memmesheimer, R.-m. (2015). A unified dynamic model for learning, replay, and sharp-wave / ripples. *Journal of Neuroscience*, 35(49):16236–16258.

Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., and De La Rocha, J. (2017). UP-DOWN cortical dynamics reflect state transitions in a bistable network. *eLife*, 6:e22425.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10(1):100–107.

Jiang, H., Liu, S., Geng, X., Caccavano, A., Conant, K., Vicini, S., and Wu, J. (2018). Pacing hippocampal sharp-wave ripples with weak electric stimulation. *Frontiers in Neuroscience*, 12:164.

Jiang, X., Shamie, I., Doyle, W. K., Friedman, D., Dugan, P., Devinsky, O., Eskandar, E., Cash, S. S., Thesen, T., and Halgren, E. (2017). Replay of large-scale spatio-temporal patterns from waking during subsequent NREM sleep in human cortex. *Scientific Reports*, 7(1):1–17.

Jolivet, R., Kobayashi, R., Rauch, A., Naud, R., Shinomoto, S., and Gerstner, W. (2008). A benchmark test for a quantitative assessment of simple neuron models. *Journal of Neuroscience Methods*, 169(2):417–424.

Jouvet, M., Michel, F., and Courjon, J. (1959). L'activite electrique du rhinencephale au cours du sommeil chez le chat. *C R Soc Biol (Paris)*, 153:101–105.

Káli, S. and Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature Neuroscience*, 7(3):286–294.

Kametani, H. and Kawamura, H. (1990). Alterations in acetylcholine release in the rat hippocampus during sleep-wakefulness detected by intracerebral dialysis. *Life Sciences*, 47(5):421–426.

Kamondi, A., Acsády, L., Wang, X. J., and Buzsáki, G. (1998). Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: Activity-dependent phase-precession of action potentials. *Hippocampus*, 8(3):244–261.

Karlsson, M. P. and Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, 12(7):913–918.

Karnani, X. M. M., Jackson, J., Ayzenshtat, I., Sichani, X. A. H., Manoocheri, K., Kim, S., and Yuste, R. (2016). Opening holes in the blanket of inhibition: localized lateral disinhibition by VIP interneurons. *Journal of Neuroscience*, 36(12):3471–3480.

Katona, L., Lapray, D., Viney, T. J., Oulhaj, A., Borhegyi, Z., Micklem, B. R., Klausberger, T., and Somogyi, P. (2014). Sleep and movement differentiates actions of two types of somatostatin-expressing GABAergic interneuron in rat hippocampus. *Neuron*, 82(4):872–886.

Katona, L., Micklem, B., Borhegyi, Z., Swiejkowski, D. A., Valenti, O., Viney, T. J., Kotzadimitriou, D., Klausberger, T., and Somogyi, P. (2017). Behavior-dependent activity patterns of GABAergic long-range projecting neurons in the rat hippocampus. *Hippocampus*, 27(4):359–377.

Klausberger, T., Magill, P., Marton, L., Roberts, J., Cobden, P., Buzsáki, G., and Somogyi, P. (2003). Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature*, 421(6925):844–848.

Klausberger, T., Marton LF, O'Neill, J., Huck, J. H. J., Dalezios, Y., Fuentealba, P., Suen, W. Y., Papp, E., Kaneko, T., Watanabe, M., Csicsvari, J., and Somogyi, P. (2005). Complementary roles of cholecystokinin- and parvalbumin-expressing GABAergic neurons in hippocampal network oscillations. *Journal of Neuroscience*, 25(42):9782–9793.

Klausberger, T. and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science*, 321(5885):53–57.

Kneisler, T. B. and Dingledine, R. (1995). Synaptic input from CA3 pyramidal cells to dentate basket cells in rat hippocampus. *Journal of Physiology*, 487(1):125–146.

Kobayashi, R., Tsubo, Y., and Shinomoto, S. (2009). Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Frontiers in Computational Neuroscience*, 3(9):1–11.

Kohus, Z., Káli, S., Rovira, L., Schlingloff, D., Papp, O., Freund, T. F., Hájos, N., and Gulyás, A. I. (2016). Properties and dynamics of inhibitory synaptic communication within the CA3 microcircuits of pyramidal cells and interneurons expressing parvalbumin or cholecystokinin. *Journal of Physiology*, 82(6):1496–1514.

Koniaris, E., Drimala, P., Sotiriou, E., and Papatheodoropoulos, C. (2011). Different effects of zolpidem and diazepam on hippocampal sharp wave-ripple activity in vitro. *Neuroscience*, 175:224–234.

Kraus, B. J., Robinson, R. J., White, J. A., Eichenbaum, H., and Hasselmo, M. E. (2013). Hippocampal "time cells": time versus path integration. *Neuron*, 78(6):1090–1101.

Kraushaar, U. and Jonas, P. (2000). Efficacy and stability of quantal GABA release at a hippocampal nnterneuron-principal neuron Synapse. *Journal of Neuroscience*, 20(15):5594–5607.

Kubota, D., Colgin, L. L., Casale, M., Brucher, F. A., and Lynch, G. (2003). Endogenous waves in hippocampal slices. *Journal of Neurophysiology*, 89(1):81–9.

Kudrimoti, H. S., Barnes, C. A., and McNaughton, B. L. (1999). Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, 19(10):4090–4101.

Kuki, T., Fujihara, K., Miwa, H., Tamamaki, N., Yanagawa, Y., and Mushiake, H. (2015). Contribution of parvalbumin and somatostatin-expressing GABAergic neurons to slow oscillations and the balance in beta-gamma oscillations across cortical layers. *Frontiers in Neural Circuits*, 9:6.

Kumar, A., Schrader, S., Aertsen, A., and Rotter, S. (2008). The high-conductance state of cortical networks. *Neural Computation*, 20(1):1–43.

Kwan, A. C. and Dan, Y. (2012). Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Current Biology*, 22(16):1459–1467.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N. J. M. A., McNaughton, B. L., and Pennartz, C. M. A. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *Journal of Neuroscience*, 28(25):6372–6382.

Lapicque, L. (1907). Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *Journal de Physiologie et de Pathologie Generale*, 9:620–635.

Lapray, D., Lasztoczi, B., Lagler, M., Viney, T. J., Katona, L., Valenti, O., Hartwich, K., Borhegyi, Z., Somogyi, P., and Klausberger, T. (2012). Behavior-dependent specialization of identified hippocampal interneurons. *Nature Neuroscience*, 15(9):1265–1271.

Lasztoczi, B., Tukker, J. J., Somogyi, P., and Klausberger, T. (2011). Terminal field and firing selectivity of cholecystokinin-expressing interneurons in the hippocampal CA3 area. *Journal of Neuroscience*, 31(49):18073–18093.

Latchoumane, C. F. V., Ngo, H. V. V., Born, J., and Shin, H. S. (2017). Thalamic spindles promote memory formation during sleep through triple phase-locking of cortical, thalamic, and hippocampal rhythms. *Neuron*, 95(2):424–435.

Le Van Quyen, M., Bragin, A., Staba, R., Crepon, B., Wilson, C. L., and Engel, J. (2008). Cell type-specific firing during ripple oscillations in the hippocampal formation of humans. *Journal of Neuroscience*, 28(24):6104–6110.

Lee, A. K. and Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194.

Lee, S. H., Marchionni, I., Bezaire, M., Varga, C., Danielson, N., Lovett-Barron, M., Losonczy, A., and Soltesz, I. (2014). Parvalbumin-positive basket cells differentiate among hippocampal pyramidal cells. *Neuron*, 82(5):1129–1144.

Leinekugel, X., Khazipov, R., Cannon, R., Hirase, H., Ben-Ari, Y., and Buzsáki, G. (2002). Correlated bursts of activity in the neonatal hippocampus in vivo. *Science*, 296(5575):2049–2052.

Letzkus, J. J., Wolff, S. B., and Lüthi, A. (2015). Disinhibition, a circuit mechanism for associative learning and memory. *Neuron*, 88(2):264–276.

Leung, L. S. and Yu, H.-W. (2017). Theta-frequency resonance in hippocampal CA1 neurons in vitro demonstrated by sinusoidal current injection. *Journal of Neurophysiology*, 79(3):1592–1596.

Levenstein, D., Buzsáki, G., and Rinzel, J. (2019). NREM sleep in the rodent neocortex and hippocampus reflects excitable dynamics. *Nature Communications*, 10(1):2478.

Levy, W. B. and Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–797.

Li, C. Y. T., Poo, M. M., and Dan, Y. (2009). Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324(5927):643–646.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, 484(7394):381.

Llorens-Martin, M., Blazquez-Llorca, L., Benavides-Piccione, R., Rabano, A., Hernandez, F., Avila, J., and DeFelipe, J. (2014). Selective alterations of neurons and circuits related to early memory loss in Alzheimer's disease. *Frontiers in Neuroanatomy*, 8:38.

Losonczy, A., Zhang, L., Shigemoto, R., Somogyi, P., and Nusser, Z. (2002). Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurones. *Journal of Physiology*, 542(1):193–210.

Luo, W. and Guan, J.-S. (2018). Do brain oscillations orchestrate memory? *Brain Science Advances*, 4(1):16–33.

Luz, Y. and Shamir, M. (2012). Balancing feed-forward excitation and inhibition via hebbian inhibitory synaptic plasticity. *PLoS Computational Biology*, 8(1):e1002334.

Lynch, G. S., Dunwiddie, T., and Gribkoff, V. (1977). Heterosynaptic depression: A postsynaptic correlate of long-term potentiation. *Nature*, 266(5604):737–739.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749.

MacVicar, B. A. and Dudek, F. E. (1980). Local synaptic circuits in rat hippocampus: interactions between pyramidal cells. *Brain Research*, 184(1):220–223.

Maier, N., Güldenagel, M., Söhl, G., Siegmund, H., Willecke, K., and Draguhn, A. (2002). Reduction of high-frequency network oscillations (ripples) and pathological network discharges in hippocampal slices from connexin 36-deficient mice. *Journal of Physiology*, 541(2):521–528.

Maier, N. and Kempter, R. (2017). Hippocampal sharp wave/ripple complexes - Physiology and mechanisms. In *Cognitive Neuroscience of Memory Consolidation*, pages 227–249. Springer.

Maier, N., Morris, G., Johenning, F. W., and Schmitz, D. (2009). An approach for reliably investigating hippocampal sharp wave-ripples in vitro. *PLoS ONE*, 4(9):e6925.

Maier, N., Nimmrich, V., and Draguhn, A. (2003). Cellular and network mechanisms underlying spontaneous sharp wave-ripple complexes in mouse hippocampal slices. *Journal of Physiology*, 550(3):873–887.

Maier, N., Tejero-Cantero, A., Dorrn, A. L., Winterer, J., Beed, P. S., Morris, G., Kempter, R., Poulet, J. F. a., Leibold, C., and Schmitz, D. (2011). Coherent phasic excitation during hippocampal ripples. *Neuron*, 72(1):137–152.

Malerba, P. and Bazhenov, M. (2019). Circuit mechanisms of hippocampal reactivation during sleep. *Neurobiology of Learning and Memory*, 160:98–107.

Malerba, P., Krishnan, G. P., Fellous, J. M., and Bazhenov, M. (2016). Hippocampal CA1 ripples as inhibitory transients. *PLoS Computational Biology*, 12(4):1–30.

Maloney, K. J., Cape, E. G., Gotman, J., and Jones, B. E. (1997). High-frequency gamma electroencephalogram activity in association with sleep-wake states and spontaneous behaviors in the rat. *Neuroscience*, 76(2):541–555.

Manns, J. R. and Eichenbaum, H. (2006). Evolution of declarative memory. *Hippocampus*, 16(9):795–808.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 262:23–81.

Marrosu, F., Portas, C., Mascia, M. S., Casu, M. A., Fà, M., Giagheddu, M., Imperato, A., and Gessa, G. L. (1995). Microdialysis measurement of cortical and hippocampal acetylcholine release during sleep-wake cycle in freely moving cats. *Brain Research*, 671(2):329–332.

Marshall, L. and Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, 11(10):442–450.

Marshall, L., Helgadóttir, H., Mölle, M., and Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature*, 444(7119):610–613.

Martin, S. J., Grimwood, P. D., and Morris, R. G. M. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, 23(1):649–711.

Martinez, J. L. J. and Kesner, R. P. (2014). *Learning and memory: A biological view*. Elsevier.

Mayford, M., Siegelbaum, S. A., and Kandel, E. R. (2012). Synapses and memory storage. *Cold Spring Harbor Perspectives in Biology*, 4(6):1–18.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457.

Mejias, J. F., Kappen, H. J., and Torres, J. J. (2010). Irregular dynamics in up and down cortical states. *PLoS ONE*, 5(11):e13651.

Memmesheimer, R.-M. (2010). Quantitative prediction of intermittent high-frequency oscillations in neural networks with supralinear dendritic interactions. *Proceedings of the National Academy of Sciences*, 107(24):11092–11097.

Menendez de La Prida, L., Huberfeld, G., Cohen, I., and Miles, R. (2006). Threshold behavior in the initiation of hippocampal population bursts. *Neuron*, 49(1):131–142.

Merker, B. (2013). Cortical gamma oscillations: the functional key is activation, not cognition. *Neuroscience and biobehavioral reviews*, 37(3):401–17.

Miles, R. and Wong, R. K. (1986). Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus. *Journal of Physiology*, 373(1):397–418.

Mizuseki, K. and Buzsáki, G. (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell Reports*, 4(5):1010–1021.

Mölle, M., Eschenko, O., Gais, S., Sara, S. J., and Born, J. (2009). The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. *European Journal of Neuroscience*, 29(5):1071–1081.

Mölle, M., Yeshenko, O., Marshall, L., Sara, S. J., and Born, J. (2006). Hippocampal sharp wave-ripples linked to slow oscillations in rat slow-wave sleep. *Journal of Neurophysiology*, 96(1):62–70.

Monmaur, P., Collet, A., Puma, C., Frankel-Kohn, L., and Sharif, A. (1996). Relations between acetylcholine release and electrophysiological characteristics of theta rhythm: A microdialysis study in the urethane- anesthetized rat hippocampus. *Brain Research Bulletin*, 42(2):141–146.

Montbrió, E., Pazó, D., and Roxin, A. (2015). Macroscopic description for networks of spiking neurons. *Physical Review X*, 5(2):021028.

Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology*, 98(3):1125–1139.

Morris, R. G., Anderson, E., Lynch, G. S., and Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature*, 319(6056):774–776.

Müller, G. E. and Pilzecker, A. (1990). *Experimentelle Beiträge zur Lehre vom Gedächtnis*. JA Barth.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., and Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(21):9497–9507.

Nadel, L. and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7(2):217–227.

Nakashiba, T., Buhl, D. L., McHugh, T. J., and Tonegawa, S. (2009). Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. *Neuron*, 62(6):781–787.

Nakazawa, K., McHugh, T. J., Wilson, M. A., and Tonegawa, S. (2004). NMDA receptors, place cells and hippocampal spatial memory. *Nature Reviews Neuroscience*, 5(5):361–372.

Neske, G. T., Patrick, S. L., and Connors, B. W. (2015). Contributions of diverse excitatory and inhibitory neurons to recurrent network activity in cerebral cortex. *Journal of Neuroscience*, 35(3):1089–1105.

Neves, G., Cooke, S., and Bliss, T. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature Reviews Neuroscience*, 9(1):65–75.

Ngo, H. V. V., Martinetz, T., Born, J., and Mölle, M. (2013). Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron*, 78(3):545–553.

Nguyen, D. P. (2009). Characterizing the frequency structure of fast oscillations in the rodent hippocampus. *Frontiers in Integrative Neuroscience*, 3(11):1–14.

Niethard, N., Ngo, H.-V. V., Ehrlich, I., and Born, J. (2018). Cortical circuit activity underlying sleep slow oscillations and spindles. *Proceedings of the National Academy of Sciences*, 115(39):E9220–E9229.

Nimmrich, V., Maier, N., Schmitz, D., and Draguhn, A. (2005). Induced sharp wave-ripple complexes in the absence of synaptic inhibition in mouse hippocampal slices. *Journal of Physiology*, 563(3):663–670.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175.

O'Keefe, J. and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330.

Okun, M. and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5):535–537.

Oliva, A., Fernández-Ruiz, A., Buzsáki, G., and Berényi, A. (2016). Role of hippocampal CA2 region in triggering sharp-wave ripples. *Neuron*, 91(6):1342–1355.

Omura, Y., Carvalho, M. M., Inokuchi, K., and Fukai, T. (2015). A lognormal recurrent network model for burst generation during hippocampal sharp waves. *Journal of Neuroscience*, 35(43):14585–14601.

O'Neill, J., Boccara, C., Stella, F., Schönenberger, P., and Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. *Science*, 355(6321):184–188.

O'Neill, J., Senior, T. J., Allen, K., Huxter, J. R., and Csicsvari, J. (2008). Reactivation of experience-dependent cell assembly patterns in the hippocampus. *Nature Neuroscience*, 11(2):209–215.

Ostojic, S. and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. *PLoS Computational Biology*, 7(1):e1001056.

Ozen, S., Sirota, A., Belluscio, M. A., Anastassiou, C. A., Stark, E., Koch, C., and Buzsaki, G. (2010). Transcranial electric stimulation entrains cortical neuronal populations in rats. *Journal of Neuroscience*, 30(34):11476–11485.

Pala, A. and Petersen, C. C. H. (2015). In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission in layer 2/3 mouse barrel cortex. *Neuron*, 85(1):68–76.

Pangalos, M., Donoso, J. R., Winterer, J., Zivkovic, A. R., Kempter, R., Maier, N., and Schmitz, D. (2013). Recruitment of oriens-lacunosum-moleculare interneurons during hippocampal ripples. *Proceedings of the National Academy of Sciences*, 110(11):4398–403.

Papatheodoropoulos, C. (2010). Patterned activation of hippocampal network (∼10 Hz) during in vitro sharp wave-ripples. *Neuroscience*, 168(2):429–442.

Papatheodoropoulos, C. and Kostopoulos, G. (2002). Spontaneous, low frequency (∼2-3 Hz) field activity generated in rat ventral hippocampal slices perfused with normal medium. *Brain Research Bulletin*, 57(2):187–193.

Papatheodoropoulos, C., Sotiriou, E., Kotzadimitriou, D., and Drimala, P. (2007). At clinically relevant concentrations the anaesthetic/amnesic thiopental but not the anticonvulsant phenobarbital interferes with hippocampal sharp wave-ripple complexes. *BMC Neuroscience*, 8(1):60.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894):1322–1327.

Patel, J., Schomburg, E. W., Berényi, A., Fujisawa, S., and Buzsáki, G. (2013). Local generation and propagation of ripples along the septotemporal axis of the hippocampus. *Journal of Neuroscience*, 33(43):17029–41.

Pelkey, K. A., Chittajallu, R., Craig, M. T., Tricoire, L., Wester, J. C., and McBain, C. J. (2017). Hippocampal GABAergic Inhibitory Interneurons. *Physiological Reviews*, 97(4):1619–1747.

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, 108(13):5419–24.

Peyrache, A., Battaglia, F. P., and Destexhe, A. (2011). Inhibition recruitment in prefrontal cortex during sleep spindles and gating of hippocampal inputs. *Proceedings of the National Academy of Sciences*, 108(41):17207–17212.

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068–76.

Pfeiffer, B. E. and Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–9.

Ponomarenko, A. A., Korotkova, T. M., Sergeeva, O. A., and Haas, H. L. (2004). Multiple GABAA receptor subtypes regulate hippocampal ripple oscillations. *European Journal of Neuroscience*, 20(8):2141–2148.

Poo, M.-M., Martin, K. C., Gage, F. H., Shtrahman, M., Rudenko, A., Pignatelli, M., Gonçalves, J. T., Tonegawa, S., Stevens, C., Johnston, S. T., Tsai, L.-H., Bonhoeffer, T., Dan, Y., Tsien, R. W., Mullins, C., Ryan, T. J., Buzsáki, G., Fishell, G., and Long, J. (2016). What is memory? The present state of the engram. *BMC Biology*, 14(1):1–18.

Poulet, J. F. A. and Crochet, S. (2019). The cortical states of wakefulness. *Frontiers in Systems Neuroscience*, 12(64).

Povysheva, N. V., Zaitsev, A. V., Gonzalez-Burgos, G., and Lewis, D. A. (2013). Electrophysiological heterogeneity of fast-spiking interneurons: chandelier versus basket cells. *PLoS ONE*, 8(8):e70553.

Rattenborg, N. C., Martinez-Gonzalez, D., Roth, T. C., and Pravosudov, V. V. (2011). Hippocampal memory consolidation during sleep: A comparison of mammals and birds. *Biological Reviews*, 86(3):658–691.

Renart, A., Brunel, N., and Wang, X.-J. (2003). Mean-field theory of recurrent cortical networks: working memory circuits with irregularly spiking neurons. In J. Feng (Ed.), editor, *Computational Neuroscience: a comprehensive approach*, pages 432–490. London: Chapman and Hall.

Renart, A., Rocha, J. D., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–591.

Reyes, A., Lujan, R., Rozov, A., Burnashev, N., Somogyi, P., and Sakmann, B. (1998). Target-cell-specific facilitation and depression in neocortical circuits. *Nature Neuroscience*, 1(4):279–285.

Ribeiro, S., Gervasoni, D., Soares, E. S., Zhou, Y., Lin, S. C., Pantoja, J., Lavine, M., and Nicolelis, M. A. (2004). Long-lasting novelty-induced neuronal reverberation during slow-wave sleep in multiple forebrain areas. *PLoS Biology*, 2(1):0126–0137.

Richardson, M. J. (2004). Effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Physical Review E*, 69(5):051918.

Rieubland, S., Roth, A., and Häusser, M. (2014). Structured connectivity in cerebellar inhibitory networks. *Neuron*, 81(4):913–929.

Romani, S. and Tsodyks, M. (2015). Short-term plasticity based network model of place cells dynamics. *Hippocampus*, 25(1):94–105.

Roux, L., Hu, B., Eichler, R., Stark, E., and Buzsáki, G. (2017). Sharp wave ripples during learning stabilize the hippocampal spatial map. *Nature Neuroscience*, 20(6):845–853.

Royer, S., Zemelman, B. V., Losonczy, A., Kim, J., Chance, F., Magee, J. C., and Buzsáki, G. (2012). Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature Neuroscience*, 15(5):769–775.

Ryan, T. J., Roy, D. S., Pignatelli, M., Arons, A., and Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. *Science*, 348(6238):1007–1013.

Sasaki, T., Matsuki, N., and Ikegaya, Y. (2014). Interneuron firing precedes sequential activation of neuronal ensembles in hippocampal slices. *European Journal of Neuroscience*, 39(12):2027–2036.

Schlingloff, D., Káli, S., Freund, T. F., Hájos, N., and Gulyás, A. I. (2014). Mechanisms of sharp wave initiation and ripple generation. *Journal of Neuroscience*, 34(34):11385–11398.

Schmitz, D., Schuchmann, S., Fisahn, A., Draguhn, A., Buhl, E. H., Petrasch-Parwez, E., Dermietzel, R., Heinemann, U., and Traub, R. D. (2001). Axo-axonal coupling: a novel mechanism for ultrafast neuronal communication. *Neuron*, 31(5):831–840.

Schönberger, J., Draguhn, A., and Both, M. (2014). Lamina-specific contribution of glutamatergic and GABAergic potentials to hippocampal sharp wave-ripple complexes. *Frontiers in Neural Circuits*, 8(103):1–11.

Schwalger, T., Deger, M., and Gerstner, W. (2017). Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS Computational Biology*, 13(4):e1005507.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, 20(11):11–21.

Semon, R. W. (1921). *The mneme*. G. Allen & Unwin Limited.

Shein-Idelson, M., Ondracek, J. M., Liaw, H. P., Reiter, S., and Laurent, G. (2016). Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science*, 352(6285):590–595.

Shen, B. and McNaughton, B. L. (1996). Modeling the spontaneous reactivation of experience-specific hippocampal cell assembles during sleep. *Hippocampus*, 6(6):685–692.

Sheridan, R. D. and Sutor, B. (1990). Presynaptic M1 muscarinic cholinoceptors mediate inhibition of excitatory synaptic transmission in the hippocampus in vitro. *Neuroscience Letters*, 108(3):273–278.

Shimizu, E., Tang, Y. P., Rampon, C., and Tsien, J. Z. (2000). NMDA receptor-dependent synaptic reinforcement as a crucial process for memory consolidation. *Science*, 290(5494):1170–1174.

Shomali, S. R., Ahmadabadi, M. N., Shimazaki, H., and Rasuli, S. N. (2018). How does transient signaling input affect the spike timing of postsynaptic neuron near the threshold regime: an analytical study. *Journal of Computational Neuroscience*, 44(2):147–171.

Shpiro, A., Curtu, R., Rinzel, J., and Rubin, N. (2007). Dynamical characteristics common to neuronal competition models. *Journal of neurophysiology*, 97(1):462–473.

Shpiro, A., Moreno-Bote, R., Rubin, N., and Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *Journal of Computational Neuroscience*, 27(1):37–54.

Shriki, O., Hansel, D., and Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural Computation*, 15(8):1809–1841.

Siapas, A. G. and Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron*, 21(5):1123–1128.

Sieghart, W., Sieghart, W., Sperk, G., and Sperk, G. (2002). Subunit composition, distribution and function of GABA. *Current Topics in Medicinal Chemistry*, 2(8):795–816.

Sik, A., Penttonen, M., Ylinen, A., and Buzsáki, G. (1995). Hippocampal CA1 interneurons: an in vivo intracellular labeling study. *The Journal of Neuroscience*, 15(10):6651–6665.

Silberberg, G. and Markram, H. (2007). Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron*, 53(5):735–746.

Sirota, A. and Buzsáki, G. (2005). Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus and Related Systems*, 3(4):245–259.

Sirota, A., Csicsvari, J., Buhl, D., and Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences*, 100(4):2065–2069.

Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., and Buzsáki, G. (2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron*, 60(4):683–697.

Skaggs, W. E. and McNaughton, B. L. (1996). Replay of neural firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257):1870–1873.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–172.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334–50.

Somogyi, P., Katona, L., Klausberger, T., Lasztóczi, B., and Viney, T. J. (2014). Temporal redistribution of inhibition over neuronal subcellular domains underlies state-dependent rhythmic change of excitability in the hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120518.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):e68.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, 99(2):195–231.

Squire, L. R. and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, 5(2):169–177.

Staresina, B. P., Alink, A., Kriegeskorte, N., and Henson, R. N. (2013). Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences*, 110(52):21159–21164.

Stark, E., Eichler, R., Roux, L., Fujisawa, S., Rotstein, H. G., and Buzsáki, G. (2013). Inhibition-induced theta resonance in cortical circuits. *Neuron*, 80(5):1263–1276.

Stark, E., Roux, L., Eichler, R., Senzai, Y., Royer, S., and Buzsáki, G. (2014). Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron*, 83(2):467–480.

Stein, R. B. (1967). Some models of neuronal variability. *Biophysical Journal*, 7(1):37–68.

Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience*, 137(4):1087–1106.

Steriade, M. and Amzica, F. (1998). Slow sleep oscillation, rhythmic K-complexes, and their paroxysmal developments. *Journal of sleep research*, 7(Suppl 1):30–35.

Steriade, M., Contreras, D., Dossi, C., and Nunez, A. (1993a). The slow (<1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *Journal of Neuroscience*, 13(8):3284–3299.

Steriade, M., McCormick, D. A., and Sejnowski, T. J. (1993b). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 261(5119):361–364.

Steriade, M., Nunez, A., and Amzica, F. (1993c). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *Journal of Neuroscience*, 13(8):3252–3265.

Steriade, M., Nunez, A., and Amzica, F. (1993d). Intracellular analysis of relations between the slow (< 1 Hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *Journal of Neuroscience*, 13(8):3266–3283.

Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos*. CRC Press.

Sullivan, D., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., and Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. *Journal of Neuroscience*, 31(23):8605–8616.

Sullivan, D., Mizuseki, K., Sorgi, A., and Buzsáki, G. (2014). Comparison of sleep spindles and theta oscillations in the hippocampus. *Journal of Neuroscience*, 34(2):662–674.

Suzuki, S. S. and Smith, G. K. (1988). Spontaneous EEG spikes in the normal hippocampus. IV. Effects of medial septum and entorhinal cortex lesions. *Electroencephalography and Clinical Neurophysiology*, 70(1):73–83.

Swaminathan, A., Wichert, I., Schmitz, D., and Maier, N. (2018). Involvement of mossy cells in sharp wave-ripple activity in vitro. *Cell Reports*, 23(9):2541–2549.

Szabó, G. G., Holderith, N., Gulyás, A. I., Freund, T. F., and Hájos, N. (2010). Distinct synaptic properties of perisomatic inhibitory cell types and their different modulation by cholinergic receptor activation in the CA3 region of the mouse hippocampus. *European Journal of Neuroscience*, 31(12):2234–2246.

Takesian, A. E., Bogart, L. J., Lichtman, J. W., and Hensch, T. K. (2018). Inhibitory circuit gating of auditory critical-period plasticity. *Nature Neuroscience*, 21(2):218–227.

Tartaglia, E. M. and Brunel, N. (2017). Bistability and up/down state alternations in inhibition-dominated randomly connected networks of LIF neurons. *Scientific Reports*, 7(1):1–14.

Taxidis, J., Coombes, S., Mason, R., and Owen, M. R. (2012). Modeling sharp wave-ripple complexes through a CA3-CA1 network model with chemical synapses. *Hippocampus*, 22(5):995–1017.

Taxidis, J., Mizuseki, K., Mason, R., and Owen, M. R. (2013). Influence of slow oscillation on hippocampal activity and ripples through cortico-hippocampal synaptic interactions, analyzed by a cortical-CA3-CA1 network model. *Frontiers in Computational Neuroscience*, 7(3):1–19.

Timofeev, I. and Bazhenov, M. (2005). Mechanisms and biological role of thalamocortical oscillations. *Trends in Chronobiology Research*, pages 1–47.

Tonegawa, S., Morrissey, M. D., and Kitamura, T. (2018). The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, 19(8):485–498.

Traub, R. and Bibbig, A. (2000). A model of high-frequency ripples in the hippocampus based on synaptic coupling plus axon–axon gap junctions between pyramidal neurons. *Journal of Neuroscience*, 20(6):2086–2093.

Traub, R. and Schmitz, D. (2012). Axonal properties determine somatic firing in a model of in vitro CA1 hippocampal sharp wave/ripples and persistent gamma oscillations. *European Journal of Neuroscience*, 36(5):2650–2660.

Traub, R. D. and Wong, R. K. (1982). Cellular mechanism of meuronal synchronization in epilepsy. *Science*, 216(4547):745–748.

Tsien, J. Z., Huerta, P. T., and Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell*, 87(7):1327–1338.

Tukker, J. J., Lasztóczi, B., Katona, L., Roberts, J. D. B., Pissadaki, E. K., Dalezios, Y., Márton, L., Zhang, L., Klausberger, T., and Somogyi, P. (2013). Distinct dendritic arborization and in vivo firing patterns of parvalbumin-expressing basket cells in the hippocampal area CA3. *Journal of Neuroscience*, 33(16):6809–6825.

Valentino, R. and Dingledine, R. (2018). Presynaptic inhibitory effect of acetylcholine in the hippocampus. *Journal of Neuroscience*, 1(7):784–792.

Van Der Velde, F. and Kamps, M. (2001). From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience*, 13(4):479–491.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.

Van Vreeswijk, C. and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10(6):1321–1371.

Vandecasteele, M., Varga, V., Berényi, A., Papp, E., Barthó, P., Venance, L., Freund, T. F., and Buzsáki, G. (2014). Optogenetic activation of septal cholinergic neurons suppresses sharp wave ripples and enhances theta oscillations in the hippocampus. *Proceedings of the National Academy of Sciences*, 111(37):13535–13540.

Vanderwolf, C. H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and Clinical Neurophysiology*, 26(4):407–418.

Varga, C., Golshani, P., and Soltesz, I. (2012). Frequency-invariant temporal ordering of interneuronal discharges during hippocampal oscillations in awake mice. *Proceedings of the National Academy of Sciences*, 109(40):E2726–E2734.

Varga, C., Oijala, M., Lish, J., Szabo, G. G., Bezaire, M., Marchionni, I., Golshani, P., and Soltesz, I. (2014). Functional fission of parvalbumin interneuron classes during fast network events. *eLife*, 3:e04006.

Vargas, R., Thorsteinsson, H., and Karlsson, K. A. (2012). Spontaneous neural activity of the anterodorsal lobe and entopeduncular nucleus in adult zebrafish: A putative homologue of hippocampal sharp waves. *Behavioural Brain Research*, 229(1):10–20.

Viney, T. J., Lasztoczi, B., Katona, L., Crump, M. G., Tukker, J. J., Klausberger, T., and Somogyi, P. (2013). Network state-dependent inhibition of identified hippocampal CA3 axo-axonic cells in vivo. *Nature Neuroscience*, 16(12):1802–11.

Vladimirov, N., Tu, Y., and Traub, R. D. (2013). Synaptic gating at axonal branches, and sharp-wave ripples with replay: A simulation study. *European Journal of Neuroscience*, 38(10):3435–3447.

Vogels, T. P., Froemke, R. C., Doyon, N., Gilson, M., Haas, J. S., Liu, R., Maffei, A., Miller, P., Wierenga, C. J., Woodin, M. A., Zenke, F., and Sprekeler, H. (2013). Inhibitory synaptic plasticity: spike timing-dependence and putative network function. *Frontiers in Neural Circuits*, 7:119.

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062):1569–1573.

Volgushev, M., Chauvette, S., Mukovski, M., and Timofeev, I. (2006). Precise long-range synchronization of activity and silence in neocortical neurons during slow-wave sleep. *Journal of Neuroscience*, 26(21):5665–5672.

Walker, F., Möck, M., Feyerabend, M., Guy, J., Wagener, R. J., Schubert, D., Staiger, J. F., and Witte, M. (2016). Parvalbumin-and vasoactive intestinal polypeptide-expressing neocortical interneurons impose differential inhibition on Martinotti cells. *Nature Communications*, 7:13664.

Wang, D. V., Yau, H. J., Broker, C. J., Tsou, J. H., Bonci, A., and Ikemoto, S. (2015). Mesopontine median raphe regulates hippocampal ripple oscillation and memory consolidation. *Nature Neuroscience*, 18(5):728–735.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., and Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science*, 313(5790):1093–1097.

Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., and Buhl, E. H. (2000). Inhibition-based rhythms: Experimental and mathematical observations on network dynamics. *International Journal of Psychophysiology*, 38(3):315–336.

Wierenga, C. J. and Wadman, W. J. (2003). Excitatory inputs to CA1 interneurons show selective synaptic dynamics. *Journal of Neurophysiology*, 90(2):811–821.

Wierzynski, C. M., Lubenov, E. V., Gu, M., and Siapas, A. G. (2009). State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. *Neuron*, 61(4):587–596.

Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24.

Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679.

Winocur, G., Moscovitch, M., and Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia*, 48(8):2339–2356.

Witter, M. P. (1993). Organization of the entorhinal-hippocampal system: a review of current anatomical data. *Hippocampus*, 3(S1):33–44.

Witter, M. P. (2007). Intrinsic and extrinsic wiring of CA3: Indications for connectional heterogeneity. *Learning and Memory*, 14(11):705–713.

Wolff, S. B., Gründemann, J., Tovote, P., Krabbe, S., Jacobson, G. A., Müller, C., Herry, C., Ehrlich, I., Friedrich, R. W., Letzkus, J. J., and Lüthi, A. (2014). Amygdala interneuron

subtypes control fear learning through disinhibition. *Nature*, 509(7501):453–458.

Woodin, M. A., Ganguly, K., and ming Poo, M. (2003). Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. *Neuron*, 39(5):807–820.

Wu, C., Shen, H., Luk, W. P., and Zhang, L. (2002). A fundamental oscillatory state of isolated rodent hippocampus. *Journal of Physiology*, 540(2):509–527.

Xu, H., Liu, L., Tian, Y., Wang, J., Li, J., Zheng, J., Zhao, H., He, M., Xu, T.-L., Duan, S., and Xu, H. (2019). A disinhibitory microcircuit mediates conditioned social fear in the prefrontal cortex. *Neuron*, pages 1–15.

Ylinen, A., Bragin, A., and Nádasdy, Z. (1995). Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *Journal of Neuroscience*, 15(1):30–46.

Zhang, L., Sheppy, E., and Wu, C. (2006). Regulation of hippocampal sharp waves by Ca 2+-dependent slow after hyperpolarization. *Critical Reviews in Neurobiology*, 18(1-2):159–167.

Zhou, H., Neville, K. R., Goldstein, N., Kabu, S., Kausar, N., Ye, R., Nguyen, T. T., Gelwan, N., Hyman, B. T., and Gomperts, S. N. (2019). Cholinergic modulation of hippocampal calcium activity across the sleep-wake cycle. *eLife*, 8:e39777.

Zucca, S., D'Urso, G., Pasquale, V., Vecchia, D., Pica, G., Bovetti, S., Moretti, C., Varani, S., Molano-Mazón, M., Chiappalone, M., Panzeri, S., and Fellin, T. (2017). An inhibitory gate for state transition in cortex. *eLife*, 6:e26177.

Acknowledgements

As this wonderful journey comes to an end, it is time to acknowledge the people who made it possible. First, I would like to thank Richard for his inspiration, intuition, support, and patience over the course of these years. It was a pleasure to develop this project with you, and I very much enjoyed the atmosphere of freedom and continuous discovery you contributed to build.

This work also wouldn't have been the same without the support of Nikolaus. Your scientific passion and curiosity are contagious! I am also indebted to Claire and Aarti, who performed the experiments, and managed to keep up with my sometimes impatient requests, and to Dietmar, whose pragmatism helped a lot in bringing the project forward.

The time at ITB rarely felt like hard work, and I shall especially thank Tiz, Susi, and Natalie for making the atmosphere around me more interesting, motivating, and always very fun.

A big thanks also goes to Paul, for the continuous inspiration on the most diverse topics, to Caro for introducing me to the 80.000 hours project, to Martina for the useful writing advice, and different generations of ITB neuro members, who contributed to create such a nice environment, every day.

During these years, the BCCN helped a lot to connect, learn, develop, find encouragement and constant support. Being part of such a great graduate school is something I would recommend to everybody!

The ENCODS team also deserves a special mention. Thanks for the great time we spent together and the incredible conference we managed to organize. It was a really enriching and rewarding experience, which helped the development of this project more than one would think.

I would also like to thank Margret, Robert, Karin, Rike, and Andreas for the great administrative and technical assistance: it is much easier to work in an environment where you feel people are doing everything possible to support you. Thanks to Tiziano, Natalie, Christian, Nikolaus, Claire, and Dimitra for carefully revising different parts of this thesis: your advice helped shaping it in its final form.

Other important people contributed to make this thesis possible outside of the office. I wish to thank my family for the never-ending support and love, and old and new friends for making my life richer and more fun. And, finally, a special thanks to Christian for the support, affection, optimism, and encouragement. Starting the next journey with you is as exciting as completing this one.

Statutory Declaration

I declare in lieu of oath that I have written this thesis independently, without illicit assistance from third parties and using solely the aids mentioned.

Eidesstattliche Versicherung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Date: _____

Signature: _____